# Source Apportionment: Findings from the U.S. Supersites Program

**John G. Watson, L.-W. Antony Chen, Judith C. Chow, Prakash Doraiswamy, and Douglas H. Lowenthal**
*Division of Atmospheric Sciences, Desert Research Institute, Reno, NV*

## ABSTRACT

Receptor models are used to identify and quantify source contributions to particulate matter and volatile organic compounds based on measurements of many chemical components at receptor sites. These components are selected based on their consistent appearance in some source types and their absence in others. UNMIX, positive matrix factorization (PMF), and effective variance are different solutions to the chemical mass balance (CMB) receptor model equations and are implemented on available software. In their more general form, the CMB equations allow spatial, temporal, transport, and particle size profiles to be combined with chemical source profiles for improved source resolution. Although UNMIX and PMF do not use source profiles explicitly as input data, they still require measured profiles to justify their derived source factors. The U.S. Supersites Program provided advanced datasets to apply these CMB solutions in different urban areas. Still lacking are better characterization of source emissions, new methods to estimate profile changes between source and receptor, and systematic sensitivity tests of deviations from receptor model assumptions.

## INTRODUCTION

Receptor-oriented source apportionment models infer source contributions and atmospheric processes from air quality measurements. Receptor models complement, rather than replace, source-oriented dispersion and chemical transformation models that begin with source emission rates to estimate ambient concentrations.[1,2] Source and receptor models are mathematical representations of reality, requiring simplifying assumptions that create uncertainty. Applying both types of models to the same situation allows them to be improved when their results diverge and lends confidence to their results when they agree.[3]

### IMPLICATIONS

Receptor modeling is useful for the identification and quantification of source contributions for emission reduction strategies. Additional chemical measurements that mark specific sources are needed to improve the utility of existing samples. Continuous monitoring systems allow the transport direction of sources to be better specified. As with all models, receptor model assumptions must be challenged and source apportionment uncertainties assessed.

Receptor models have been used to develop emission reduction strategies for attaining total suspended particulates, $PM_{10}$, $PM_{2.5}$ (particulate matter with aerodynamic diameters less than 10 and 2.5 µg, respectively), and ozone ($O_3$) standards, decreasing human exposures to toxic substances, and improving visibility.[4–8] As part of these reductions many of the chemical markers for major sources, such as lead in gasoline engine exhaust and trace elements in primary industrial emissions, have been reduced or eliminated. After emissions from on-road engine exhaust and residential wood combustion (RWC) were quantified and controls introduced, separating these contributions from high emitters, cold starts, non-road engines, and wildfires became more important. Determining sources of particulate matter (PM)-forming gaseous precursors of secondary sulfates, nitrates, and secondary organic aerosol (SOA) is also important. Addressing these issues with receptor models requires new and more specific measurement techniques and better ways to handle the data.

Since 1999, the U.S. Supersites Program[9–18] has developed and tested new methods for monitoring particles and their aerosol precursors. During the same period, the Speciation Trends Network (STN)[19] was deployed in urban areas to complement the $PM_{2.5}$ mass, elements, ions, and carbon measurements from the non-urban Interagency Monitoring of PROtected Visual Environments (IMPROVE) network.[20] The Southeastern Aerosol Research and Characterization (SEARCH)[21] experiment was also initiated. Methods to obtain samples of source emissions relevant to ambient particle and gas characteristics were combined with new laboratory methods to identify markers and with aerosol evolution models to determine how these characteristics change soon after emission. These have advanced the ability of receptor models to better estimate source contributions to ambient concentrations.

More than 1500 articles and reports have been identified since mathematical receptor models were first applied to air quality measurements,[22] and several reviews and evaluations have been published.[23–40] This paper intends to address the following question: "How well can we identify and quantify source contributions using receptor models?" Except for historical and exemplary purposes, this paper limits its investigation to work published since 2000, with a focus on PM source apportionment in the vicinity of the Supersite cities of Atlanta, GA; Baltimore, MD; Fresno, CA; Houston, TX; Los Angeles, CA; New York, NY; Pittsburgh, PA; and St. Louis, MO.

## RECEPTOR MODEL FORMULATION

Most receptor models present solutions to the following equations:

$$C_{iklmn} = \sum_{j=1}^{J} F_{ijm} T_{ijklmn} S_{jklmn} \quad \text{for } i = 1 \text{ to } I \qquad (1)$$

The indices are defined as:

- $i$ represents a quantifiable chemical element, compound, or physical property that is expected to have markedly different proportions to other characteristics in different sources.
- $j$ represents a group of emitters with similar emission compositions that differ from the compositions of other source types.
- $k$ represents the sampling period, a part of the day, day of the week, season, period before or after a control measure has been implemented, or including a special event such as a fire or dust storm.
- $l$ represents the receptor location. Monitoring sites are usually selected to determine human exposure, but they are most useful for receptor models when they also include sites that represent different spatial scales,[41] including those near suspected pollution sources such as roadways and industries, as well as regional background sites representing mixtures from many emitters.
- $m$ represents particle size fraction, the most useful being the ultrafine (<0.1 μm), fine (PM$_{2.5}$), and coarse (2.5–10 μm, PM$_{10-2.5}$) fractions.
- $n$ represents transport direction, which can be a simple wind direction for local sources, or a more complex set of curvilinear trajectories for long-range transport.

Using these indices, the variables in eq 1 are:

- $C_{iklmn}$ is the concentration (unit of μg/m³, ng/m³, ppm, or ppb) of pollutant $i$ for time period $k$ at location $l$ corresponding to particle size range $m$ and transport direction $n$. This is the measured receptor concentration.
- $F_{ijm}$ is the fractional quantity of pollutant $i$ in source type $j$ for size range $m$ (unitless). For PM measurements, profile abundances are often normalized to mass emissions from a source in the desired size range and averaged over several source tests.
- $T_{ijklmn}$ are the changes in $F_{ijm}$ during transport from the source to the receptor.
- $S_{jklmn}$ is the contribution from source type $j$ in size range $m$ from wind sector $n$ for time period $k$ at location $l$ (μg/m³, ng/m³, ppm, or ppb). Source contributions are calculated by the receptor model.

In its most common use, eq 1 is solved for $S_{jklmn}$ using $F_{ijm}$ and $C_{iklmn}$ as input data. Uncertainties of the input data are designated as $\sigma_{Ciklmn}$ (usually determined by replicate analysis and propagation of analytical and flow rate uncertainties) and $\sigma_{Fijm}$ (usually estimated as the standard deviation of the average from several source tests). Uncertainties of the source contribution estimates (SCEs) are designated as $\sigma_{Sjklmn}$ and are estimated by error propagation or by Monte Carlo simulation.[33]

Equation 1 reduces to the Chemical Element Balance of Hidy and Friedlander[42] for single samples taken at a single location and time period such that:

$$C_{it} = \sum_{j} F_{ij} S_{jt} + E_{it} \qquad (2)$$

$C_{it}$ is the concentration of the $i$th chemical species measured at time or location $t$. $F_{ij}$ is the fractional abundance of the $i$th species in the $j$th source type, $S_{jt}$ is the normalized contribution of the $j$th source at time and/or location $t$, and $E_{it}$ represents the error between the measured and calculated ambient concentrations. They assumed that $T_{ijklmn}$ was equal to one, meaning that the proportions of the different elements they used as source markers for southern California did not change between source and receptor. In their example, Hidy and Friedlander[42] used individual elements as sole markers for selected source types. Friedlander[43] recognized the limitations of available measurements, and conceptualized the requirements of measurement devices that would provide information detailed enough to bring the full potential of eq 1 into reality. Taking advantage of advanced measurement techniques that were only available recently, the Supersites Program demonstrated the potential of these concepts to attain reality.

Although $F_{ijm}$ most commonly represents chemical source profiles, they may also represent other properties such as size or temporal variability. The fine and coarse modes of the aerosol size distributions provide a good first cut at source apportionment, with most of the coarse material being composed of suspended dust, sea salt, and pollen, whereas most of the fine material contains carbon from combustion products and secondary species including sulfates, nitrates, and organics. When $F_{ijm}$ represents diurnal distributions, it shows traffic peaks near morning and evening rush hours, increasing concentrations at night during winter heating seasons and sharp spikes (<5 min) from nearby sources. Many Supersite studies explored the use of spatial and temporal variations to infer and/or confirm contributing sources.

## SOLUTIONS TO THE RECEPTOR MODEL EQUATIONS

Most of the receptor models used before and during the Supersites Program are solutions to eq 1, which can be derived from physical principles with simplifying assumptions.[35] Receptor models are sometimes incorrectly called "statistical" methods. The statistical distributions, often-missing data, and uncertainties of the input measurements do not conform to the rigorous assumptions required for statistical tests. This misnomer arises because the mathematical techniques used in receptor modeling are also used to determine and test statistical associations in other areas of science.

Table 1, a–d, summarizes several of the solutions that have been applied to the receptor model equations. Owing to the availability of software provided by the U.S. Environmental Protection Agency (EPA), the UNMIX,

**Table 1a.** Different receptor models used in the Supersite source apportionment studies: chemical mass balance (CMB).

| Chemical Mass Balance | | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |
| Effective Variance CMB [42,121] (Note that all models based on eq 1 or 2 are CMB equations. The term CMB used here reflects the historical solution in which source profiles are explicitly used as model input and a single sample effective variance solution is reported.) CMB software is currently distributed by EPA. The most recent version is the CMB 8.2, which is run in the Microsoft Windows system. | *Principle* Ambient chemical concentrations are expressed as the sum of products of species abundances in source emissions and source contributions (eq 1 or 2). These equations are solved for the source contribution estimates when ambient concentrations and source profiles are input. The single-sample effective variance least squares[122] is the most commonly used solution method because it incorporates uncertainties of ambient concentrations and source profiles in the estimate of source contributions and their uncertainties. This reduced to the tracer solution when it is assumed that there is one unique species for each source. Choices of source profiles should avoid collinearity, which occurs when chemical compositions of various source emissions are not sufficiently different.[121] *Data Needs* CMB requires source profiles, which are the mass fractions of particulate or gas species in source emissions. The species and particle size fraction measured in source emissions should match those in ambient samples to be apportioned. Several sampling and analysis methods provide time-integrated speciation of $PM_{2.5}$ and volatile organic compounds (VOCs) for CMB. Source profiles are preferably obtained in the same geographical region as the ambient samples, although using source profiles from different regions is commonly practiced in the literature. The practitioner needs to decide the source profiles and species being included in the model, on the basis of the conceptual model and model performance measures. *Output* Effective variance CMB determines, if converged, source contributions to each sample in terms of PM or VOC mass. CMB also generates various model performance measures, including correlation $R^2$, deviation $\chi^2$, residue/uncertainty ratio, and MPIN matrix that are useful for refining the model inputs to obtain the best and most meaningful source apportionment resolution. | *Strengths* <br>• Software available providing a good user interface. <br>• Provides quantitative uncertainties on source contribution estimates based on input concentrations, measurement uncertainties, and collinearity of source profiles. <br>• Quantifies contributions from source types with single particle and organic compound measurements. <br>*Weaknesses* <br>• Completely compatible source and receptor measurements are not commonly available. <br>• Assumes all observed mass is due to the sources selected in advance, which involves some subjectivity. <br>• Chemically similar sources may result in collinearity without more specific chemical markers. <br>• Typically does not apportion secondary particle constituents to sources. Must be combined with profile aging model to estimate secondary PM. |

Table 1, a–d, updated with permission from Watson and Chow.[38] Copyright Air & Waste Management Association 2004.

positive matrix factorization (PMF), and chemical mass balance (CMB) are the models most widely used in the Supersites Program. CMB solutions require measured source profiles ($F_{ijm}$) as input data. UNMIX and PMF do not use source profiles as input data, but it is untrue that source profiles are unnecessary for these models. UNMIX and PMF derive source factors that must be associated with measured source profiles to achieve meaningful source apportionment. The degree of correspondence between factor and profile should guide the selection of factors, weighting of variables, and weighting of receptor concentrations for these solutions.[44,45] Unfortunately, comparisons of UNMIX and PMF factors with measured source profiles are rarely done in publications, leading to ambiguous results.

The PMF and the pseudodeterministic receptor model (PDRM) can have an infinite number of solutions, even with non-negativity constraints. PMF provides parameters such as FPEAK and $F_{key}$ to adjust the solution on the basis of the practitioner's experience and knowledge about the sources. PDRM confines the solution within preset bounds determined from dispersion models, as described in Table 1b. Zhou et al.[46] and Ogulei et al.[47] used a multilinear engine (ME or ME2) to analyze $PM_{2.5}$ speciation data combining measurements on different time scales. This cannot be achieved by PMF without averaging measurements to the same time scale. Other constraint techniques such as partial least squares (PLS) are also used in the Supersite studies. PLS seeks to resolve factors that can explain two sets of related data simultaneously.

Certain compounds are used as markers for a source type when negligible amounts are believed to come from other sources. Examples include: elemental carbon (EC) for primary combustion emissions,[48] carbon (C) isotopes

**Table 1b.** Different receptor models used in the Supersites source apportionment studies: factor analysis.

| | Factor Analysis | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |
| PMF[123,124]<br><br>PMFx (PMF2 and PMF3) software is available from Dr. Pentti Paatero at the University of Helsinki, Finland. This software is a Microsoft DOS application. EPA distributes EPA PMF[76] version 1.1 as a Microsoft Windows application with better user interface. | *Principle*<br>PMFx contains PMF2 and PMF3. PMF2 solves the CMB equations (i.e., eqs 2 and 3) using an iterative minimization algorithm. Source profiles $F_{ij}$ and contribution $S_{jt}$ are solved simultaneously. The non-negativity constraint is implemented in the algorithm to decrease the number of possible solutions (local minimums) in the PMF analyses, because both source profile and contribution should not contain negative values. There is rotational ambiguity in all two-way factor analyses (i.e., $F_{ij}$ and $S_{jt}$ matrices may be rotated and still fit the data). PMF2 allows using the FPEAK parameter to control the rotation. A positive FPEAK value forces the program to search such solutions where there are many zeros and large values but few intermediate values in the source matrix $F_{ij}$. $F_{key}$ can further bind individual elements in $F_{ij}$ to zero. On the basis of a similar algorithm, PMF3 solves a three-way problem.<br>PMFx and UNMIX estimate $F_{ij}$ and $S_{jt}$ by minimizing:<br><br>$$Q \text{ or } \chi^2 = \sum_i \sum_t [E_{it}/\sigma_{it}]^2 = \sum_i \sum_t [(C_{it} - \sum_j F_{ij}S_{jt})/\sigma_{it}]^2 \qquad (3)$$<br><br>Where the weighing factor, $\sigma_{it}$, represents the magnitude of $E_{it}$, PMFx limits solutions of eq 2 to non-negative $F_{ij}$ and $S_{jt}$.<br>*Data Needs*<br>A large number of ambient samples (usually much more than the number of factors in the model) are required to produce a meaningful solution. Species commonly used in PMF are also those in CMB. Weighting factors associated with each measurement need to be assigned before analysis. The practitioner also needs to decide the number of factors, FPEAK, and $F_{key}$ in the model.<br>*Output*<br>PMFx reports all the elements in $F_{ij}$ and $S_{jt}$ matrices (PMF2). It also calculates model performance measures such as deviation $\chi^2$ and standard deviation of each matrix element. The practitioner needs to interpret the results linking them to source profiles and source contributions. | *Strengths*<br>● Software available.<br>● Can handle missing or below-detection-limit data.<br>● Weights species concentrations by their analytical precisions.<br>● Downweight outliers in the robust mode.<br>● Derives source profiles from ambient measurements as they would appear at the receptor (does not require source measurements).<br><br>*Weaknesses*<br>● Requires large (>100) ambient datasets.<br>● Need to determine the number of retaining factors.<br>● Requires knowledge of source profiles or existing profiles to verify the representativeness of calculated factor profiles and uncertainties of factor contributions.<br>● Relies on many parameters/initial conditions adjustable to model input; sensitive to the preset parameters. |
| ME2[125]<br><br>ME2 code is available from Dr. Pentti Paatero at the University of Helsinki, Finland as a Microsoft DOS application. | *Principle*<br>The PMFx algorithm is derived from ME2. Unlike PMFx that is limited to questions in the form of eq 1 or 2, ME2 solves all models in which the data values are fitted by sums of products of unknown (and known) factor elements. The first part of the algorithm interprets instructions from the user and generates a table that specifies the model. The second part solves the model using an iterative minimization approach. Additional constraints could be programmed into the model to reduce the ambiguity in source apportionment. These constraints may include known source profiles and/or contributions (e.g., contributions are known to be zero in some cases).<br>*Data Needs*<br>Data needs are similar to those of PMFx but are more flexible. In theory, any measured or unknown variables may be included in the model as long as they satisfy linear relationships. The users need to specify the model structure, the input, and the output.<br>*Output*<br>ME2 calculates and reports all unknown variables in the model. | *Strengths*<br>● Software available.<br>● Can handle user-specified models.<br>● Possibility to include all measured variables into the model, such as speciated concentration over different time scales, size distributions, meteorological variables, and noise parameters.<br><br>*Weaknesses*<br>● Require substantial training to access the full feature of the software and develop a model.<br>● Generally requires large ambient datasets.<br>● Need to assume linear relationships between all variables.<br>● Relies on many parameters/initial conditions adjustable to model input; sensitive to the preset parameters. |

Table 1b. (cont.)

| | Factor Analysis | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |
| UNMIX[29,44,126]<br><br>UNMIX code is available from Dr. Ron Henry at the University of Southern California as an MatLab application. A stand-alone version (UNMIX version 6) is also available from EPA. | *Principle*<br>UNMIX views each sample as a data point in a multidimensional space with each dimension representing a measured species. UNMIX solves eqs 2 and 3 by using a principle component analysis (PCA) approach to reduce the number of dimensions in the space to the number of factors that produce the data, followed by an unique "edge detection" technique to identify "edges" defined by the data points in the space of reduced dimension (e.g., Figures 1 and 3). The number of factors is estimated by the NUMFACT algorithm in advance[127], which reports the $R^2$ and signal-to-noise (S/N) ratio associated with the first N principle components (PCs) in the data matrix. The number of factors should coincide with the number of PCs with S/N ratio $> 2$. Once the data are plotted on the reduced space, an edge is actually a hyperplan that signifies missing or small contribution from one or more factors. Therefore, UNMIX searches all the edges and uses them to calculate the vertices of the simplex, which are then converted back to source composition and contributions. Geometrical concepts of self-modeling curve resolution are used to ensure that the results obey (to within error) non-negativity constraints on source compositions and contributions.<br><br>*Data Needs*<br>A large number of ambient samples (usually much more than the number of factors in the model) are required to achieve a meaningful solution. Species commonly used in UNMIX are also those in CMB. The measurement precision is not required. The practitioner needs to specify the number of factors on the basis of the NUMFACT results.<br>*Output*<br>UNMIX determines all the elements in the factor ($F_{ij}$) and contribution ($S_{jt}$) matrices. It also calculates the uncertainty associated with the factor elements and model performance measures including: (1) $R^2$, (2) S/N ratio, and (3) strength. | *Strengths*<br>● Software available with graphical user interface.<br>● Does not require source measurements.<br>● Provide graphical problem diagnostic tools (e.g., species scatter plot).<br>● Provide evaluation tools (e.g., $R^2$, S/N ratio).<br>*Weaknesses*<br>● Requires large ($>100$) ambient datasets.<br>● Need to assume or predetermine number of retained factors.<br>● Does not make explicit use of errors or uncertainties in ambient measurements.<br>● Cannot use samples containing missing data in any species.<br>● Limited to a maximum of 7 or 14 (UNMIX version 6) factors.<br>● Can report multiple or no solutions.<br>● Requires knowledge of existing source profiles to evaluate the solutions. |
| PDRM[97]<br><br>PDRM was developed under the Supersites Program and requires MatLab or equivalent software to perform the calculation. | *Principle*<br>PDRM estimates contributions from selected stationary sources for a receptor site using high time-resolution measurements and meteorological data. In PDRM, eq 2 is modified to<br><br>$$C_{it} = \sum_j ER_{i,j}\left(\frac{\chi}{Q}\right)_{j,t} + E_{it} \qquad (4)$$<br><br>where $ER_{i,j}$ is interpreted as the emission rate of species $i$ from stationary source $j$ and $(\chi/Q)_{j,t}$ is the meteorological dispersion factor averaged over the time interval $t$. Equation 4 is solved for $ER_{i,j}$ and $(\chi/Q)_{j,t}$ simultaneously by a nonlinear fit minimizing the objective function, *FUN*:<br><br>$$FUN = \sum_{i=1}^{l}\sum_{t=1}^{m}\sum_{j=1}^{n}\left[ER_{i,j}\left(\frac{\chi}{Q}\right)_{j,t}^{PDRM} - C_{it}\right]^2 \qquad (5)$$<br><br>Because the number of solutions for a product of unknowns is infinite, additional constraints are set up for $(\chi/Q)_{j,t}$ on the basis of the Gaussian plume model, thus:<br><br>$$LB\left(\frac{\chi}{Q}\right)_{j,t}^{Met} \leq \left(\frac{\chi}{Q}\right)_{j,t}^{PDRM} \leq UB\left(\frac{\chi}{Q}\right)_{j,t}^{Met} \qquad (6)$$<br><br>$$\left(\frac{\chi}{Q}\right)_{j,t}^{Met} = \frac{1}{2\pi\sigma_y\sigma_z u}\exp\left(-\frac{1}{2}\frac{y^2}{\sigma_y^2}\right)\left\{\exp\left[-\frac{1}{2}\left(\frac{z-h}{\sigma_z}\right)^2\right]\right.$$<br>$$\left. + \exp\left[-\frac{1}{2}\left(\frac{z+h}{\sigma_z}\right)^2\right]\right\} \qquad (7)$$ | *Strengths*<br>● Explicitly include meteorological information and stack configuration of stationary sources into the model.<br>● Do not require source measurements.<br>● Do not need to interpret the relations between factors and sources.<br>● Commercial software (e.g., MatLab) available for performing nonlinear fit.<br>● Suitable for high time-resolution measurement.<br>*Weaknesses*<br>● Can only handle stationary sources but not area or mobile sources.<br>● Need to assume that only stationary sources are considered in the model contribute significantly for a measurement at the receptor site.<br>● Do not account for uncertainty in the measurement.<br>● Meteorological data may not be always available or accurate.<br>● Gaussian plume model may not be representative of the actual atmospheric dispersion.<br>● Sensitive to the imposed constraints (*UB* and *LB*). |

**Table 1b. (cont.)**

| Factor Analysis | | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |

|  | Equations 6 and 7 limit the solution of eq 5 within the lower (*LB*) and upper (*UB*) bound of those predicted by the Gaussian plume model using different parameterizations. |  |
|  | *Data Needs* |  |
|  | PDRM requires speciated measurements at a higher time-resolution than typical CMB or PMF applications because of the fast-changing meteorological parameters. PDRM also requires data for eq 7: transport speed ($u$), lateral and vertical dispersion parameters ($\sigma_y$ and $\sigma_z$), and stack height ($h$). |  |
|  | *Output* |  |
|  | PDRM determines emission rates and contributions from each point source considered in the model at the same time resolution as the measurement. |  |

PLS[128]

*Principle*

PLS examines the relationships between a set of predictor (independent) and response (dependent) variables. It assumes that the predictor and response variables are controlled by independent "latent variables" less in number than either the predictor or the response variables. In recent applications,[96] PM chemical composition and size distribution are used as predictor ($X$) and response ($Y$) variables, respectively. Equation 2 is modified to:

$$X_{it} = \sum_{j} T_{ij} P_{jt} + E_{it} \tag{8}$$

$$Y_{it} = \sum_{j} U_{ij} C_{jt} + D_{it} \tag{9}$$

where $T$ and $U$ are matrices of so-called "latent variables," and $P$ and $C$ are loading matrices. If $X$ and $Y$ are correlated to some degree, $T$ and $U$ would show some similarity. Equations 8 and 9 are solved by an iterative algorithm "NIPALS," which attempts to minimize $E,D,$ and the difference between $T$ and $U$ simultaneously. If $T$ and $U$ end up being close enough, the $X$ and $Y$ variables can be explained by the same latent variables. These latent variables may then be interpreted as source or source categories.

*Data Needs*

Typical applications of PLS require both chemical speciated and size-segregated measurements. The practitioner needs to decide the number of latent variables on the basis of the correlation of resulting $T$ and $U$ matrices.

*Output*

PLS calculates latent variables, which are common factors best explaining the predictor and response variables, and the residues from fitting. $R_x$ and $R_y$,

$$R_x = 1 - \text{var}(E)/\text{var}(X) \tag{10}$$

$$R_y = 1 - \text{var}(D)/\text{var}(Y) \tag{11}$$

indicate the degree to which variables $X$ and $Y$ are explained by the latent variables.

*Strengths*
- Fit two types of measurements (e.g., chemistry and size) with common factors. Provide more information to identify sources.
- Analyze strongly collinear and noisy dataset.
- Do not require source measurements.

*Weaknesses*
- Requires large (>100) ambient datasets.
- Difficult to relate latent variables to any physical quantities.
- Do not provide quantitative source contribution estimates.
- Need to decide the number of latent variables.
- Do not explicitly make use of measurement uncertainties.
- Can result in no solution.

**Table 1c.** Different receptor models used in the Supersites source apportionment studies: tracer-based method.

| Tracer-Based Method | | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |
| EF[129,130]<br>The EF method may use a MLR algorithm, which is available in most statistical and spreadsheet software | *Principle*<br>A tracer (or marker) for a particular source or source category is a species enriched heavily in the source emission against other species and other sources. Using EFs, concentration of the $i^{th}$ pollutant at a receptor site at time $t$ (i.e., $C_{i,t}$) can be expressed as:<br><br>$$C_{i,t} = \sum_j \frac{1}{EF_{i,pj}} C_{pj,t} + Z_{i,t} = \sum_j \left(\frac{F_{ij}}{F_{pj}}\right) C_{pj,t} + Z_{i,t} \qquad (12)$$<br><br>where the enrichment factor $EF_{i,pj}$ is the ratio of emission rate of the pollutant of interest ($F_{ij}$) and tracer species ($F_{pj}$) from source $j$. $C_{pj,t}$ is the concentration of tracer species for source $j$ at time $t$, and $Z_{i,t}$ represents contributions from all other sources (including the background level). The solution for eq 12 is situation-dependent. $EF_{i,pj}$ is usually unknown but may be estimated from source profiles, edges of a two-way scatter plot (e.g., Figures 1 and 3), or the ratio of $C_{i,t}$ to $C_{pj,t}$ for a particular period when it is believed that a single source is dominant. In cases where $Z_{i,t}$ is a constant, $EF_{i,pj}$ may be derived from MLR.<br>*Data Needs*<br>The minimum data needs include concentrations of all primary tracers at the receptor site. Known *EF*s or background levels are helpful.<br>*Output*<br>The EF method determines contributions to species $i$ from each source considered in the model. | *Strengths*<br>● No special software needed.<br>● Indicate presence or absence of particular emitters.<br>● Provides evidence of secondary PM formation and changes in source impacts by changes in ambient composition.<br>● Could use a large (>100) dataset or a small (e.g., <10) dataset.<br>*Weaknesses*<br>● Semiquantitative method, not specific especially when the EFs are unknown in advance.<br>● Limited to sources with unique markers.<br>● Tracer species must be exclusively from the sources or source categories examined.<br>● Provide very limited error estimates.<br>● More useful for source/process identification than for quantification. |
| NNLS[131,132]<br>The MatLab Optimization Toolbox provides a function "lsqnonneg" for performing the NNLS calculation. | *Principle*<br>NNLS also solves the EF equation (eq 12 or equivalent) with known target species and tracer concentrations. Conventional MLR solutions to eq 12 may lead to negative EFs due to the uncertainty in measurements or colinearity in source contributions. This is avoided in the NNLS approach since additional non-negative constraints are built into the algorithm, i.e.:<br><br>$$EF_{i,pj} \geqq 0 \qquad (13)$$<br><br>Utilizing orthogonal decomposition, a NNLS problem can be reduced to the more familiar least-distance programming and solved by a set of iterative subroutines developed and tested by Lawson and Hanson.[131] In a more general sense, NNLS linearly relates a response variable with a set of independent variables with only non-negative coefficients.<br>*Data Needs*<br>When applied to EF or MLR problems, NNLS requires the concentration of target (response) and tracer (independent) species.<br>*Output*<br>NNLS generates non-negative regression coefficients for an EF/MLR problem and these coefficients can be related to the source contributions. | *Strengths*<br>● Implemented by many statistical software packages.<br>● Generate only non-negative *EF*s or regression coefficients.<br>● Do not require source measurements.<br>● Possible to include meteorological or other (besides chemistry) data into the model.<br>*Weaknesses*<br>● Require a large (>100) set of ambient measurements.<br>● Semiquantitative method, not specific.<br>● Do not explicitly consider measurement uncertainties.<br>● Tracer species must be exclusively from the sources or source categories examined.<br>● Non-negative constraints may not be appropriate in some cases. |
| FAC[111] | *Principle*<br>FAC provides a simple mean of estimating the SOA production rate using the emission inventories of primary precursor VOCs. FAC is actually a source-oriented modeling technique but it does not take into account all the atmospheric processes. FAC is defined as the fraction of SOA that would result from the reactions of a particular VOC:<br><br>$$[SOA] = \sum_i FAC_i \times ([VOC_i]_0 \times \text{Fraction of VOC } i \text{ reacted}) \qquad (14)$$<br><br>where $[VOC_i]_0$ is the emission rate of $VOC_i$ and $[SOA]$ is the formation rate of SOA. Equation 14 can be viewed as an extension of eq 12 but concentrations are replaced with emission rates and EFs are replaced with *FAC*s. *FAC* and the fraction of VOC reacted under typical ambient conditions have been developed for a large number of hydrocarbons $> C_6$.[111] The most significant SOA precursors are aromatic compounds (especially toluene, xylene, and trimethylbenzenes) and terpenes. In most applications, these FACs are used directly to estimate SOA.<br>*Data Needs*<br>FAC requires the VOC emission inventory in the region of interest. The knowledge of $O_3$ and radiation intensity is also helpful for slight modifications of the FACs.<br>*Output*<br>FAC method estimates the total production rate of SOA. | *Strengths*<br>● Link SOA to primary VOC emissions so that SOA can also be treated as primary particles in the PM modeling.<br>● Simple and inexpensive.<br>*Weaknesses*<br>● Ignore the influence of aerosol concentration and temperature-dependent gas-particle partitioning on SOA yield.<br>● Limited by the accuracy of VOC emission inventory.<br>● Do not directly infer the contribution of each source to ambient SOA concentration.<br>● Difficult to verify. |

**Table 1d.** Different receptor models used in the Supersites source apportionment studies: meteorology-based method.

| | Meteorology-Based Method | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |
| CPF[134,135] | *Principle*<br>CPF estimates the probability that a given source contribution from a given wind direction will exceed a predetermined threshold criterion (e.g., upper 25th percentile of the fractional contribution from the source of interest). The calculation of CPF uses source contributions (*i.e.*, $S_{jt}$ in eq 2) determined for the receptor site and local wind direction data matching each of the source contributions in time. These data are then segregated to several sectors according to wind direction and the desired resolution (usually 36 sectors at a 10° resolution). Data with very low wind speed (e.g., < 0.1 m/sec) are usually excluded from analysis because of the uncertain wind direction. CPF is then determined by:<br><br>$$CPF(\theta) = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \quad (15)$$<br><br>where $m_{\Delta\theta}$ is the number of occurrences in the direction sector $\theta \to \theta + \Delta\theta$ that exceeds the specified threshold, and $n_{\Delta\theta}$ is the total number of wind occurrences in that sector. Because wind direction is changing rapidly, high-time resolution measurements (e.g., minutes to hours) are preferred for a CPF analysis. If the calculated source contributions represent long-term averages, wind direction needs to be averaged over the same duration. In addition to source contribution, CPF can be applied directly to pollutant concentration measurements at a receptor site.<br>*Data Needs*<br>CPF requires the time series of source contributions at a receptor site, which is usually determined by CMB or factor analysis methods using speciated measurements at the site. CPF also requires wind direction and wind speed data averaged over the same time resolution as the sampling duration.<br>*Output*<br>CPF reports the probability of "high" contribution from a particular source or factor occurring within each wind direction sector. The results are often presented in a wind rose plot (e.g., Figure 4a). | *Strengths*<br>● Infer the direction of sources or factors relative to the receptor site.<br>● Provide verification for the source identification made by factor analysis method.<br>● Easy to implement.<br>*Weaknesses*<br>● Criterion for the threshold is subjective.<br>● Absolute source contribution (or fractional contribution) may be influenced by other factors besides wind direction (e.g., wind speed, mixing height).<br>● Local and near-surface wind direction only has a limited implication for long-range transport.<br>● Easy to be biased by a small number of wind occurrences in a particular sector.<br>● Work better for stationary sources than area or mobile sources. |
| NPR[136,137] | *Principle*<br>NPR calculates the expected (averaged) source contribution as a function of wind direction following:<br><br>$$S(\theta) = \frac{\sum_i K\left(\frac{\theta - W_i}{\Delta\theta}\right) \times S_i}{\sum_i K\left(\frac{\theta - W_i}{\Delta\theta}\right)} \quad (16)$$<br><br>where $W_i$ is the wind direction for the $i^{th}$ sample and $S_i$ is the contribution from a specific source to that sample, determined from measurements at the receptor site. $K$ is a weighting function called the kernel estimator. There are many possible choices for $K$. Henry et al.[136] recommend either Gaussian or Epanechnikov functions. The most important decision in NPR is the choice of the smoothing parameter $\Delta\theta$. If $\Delta\theta$ is too large, $S(\theta)$ will be too smooth and meaningful peaks could be lost. If it is too small, $S(\theta)$ will have too many small, meaningless peaks. $\Delta\theta$ needs to be chosen according to the project-specific spatial distribution of sources. NPR also estimates the confidence intervals of $S(\theta)$ based on the asymptotic normal distribution of the kernel estimates, thus:<br><br>$$\Delta S(\theta) = \frac{\sum_i K\left(\frac{\theta - W_i}{\Delta\theta}\right) \times (S_i - S(\theta))^2}{\left(\sum_i K\left(\frac{\theta - W_i}{\Delta\theta}\right)\right)^2} \quad (17)$$ | *Strengths*<br>● Infer the direction of sources or factors relative to the receptor site.<br>● Provide verification for the source identification made by factor analysis method.<br>● Require no assumption about the function form of the relationship between wind direction and source contribution.<br>● Provide uncertainty estimates.<br>● Easy to implement.<br>*Weaknesses*<br>● Choices for the kernel estimator and smoothing factor are subjective.<br>● Absolute source contribution (or fractional contribution) may be influenced by other factors besides wind direction (e.g., wind speed, mixing height).<br>● Local and near-surface wind direction only has a limited implication for long-range transport.<br>● Easy to be biased by a small number of wind occurrences in a particular sector.<br>● Work better for stationary sources than area or mobile sources. |

**Table 1d. (cont.)**

| | Meteorology-Based Method | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |

*Data Needs*

NPR requires the same data as the CPF method, including the time series of source/factor contributions (or fractional contributions) at the receptor site and local wind direction data matching the sampling duration in time.

*Output*

NPR reports the distribution of source contribution as a function of wind direction and the confidence level associated with it.

**TSA[138]**

TSA requires the calculation of air parcel back trajectory, which is often accomplished using the HY-SPLIT model.[115,139] HY-SPLIT version 4.5 is available at *http://www.arl.noaa.gov/ready/hysplit4.html*.

*Principle*

Similar to CPF, TSA clusters the measured pollutant concentration or calculated source contribution according to the wind pattern. However, air parcel back trajectory, rather than local wind direction, is used. A back trajectory traces the air parcel backward in time from a receptor. The initial height is often between 200 and 1000 m above ground level where the wind direction could differ from the surface wind direction substantively. For each sample $i$, TSA obtains one or more trajectories and calculates their total residence time in the $j^{th}$ directional sector ($\tau_{i,j}$, i.e., the total number of 1-hr trajectory end points that fall into the sector). The pollutant concentration or source contribution in the sample, $S_i$, is then linearly apportioned into each directional sector according to $\tau_{i,j}$ and averaged over all samples to produce the directional dependent pollutant concentration/source contribution for the period of interest:

$$\overline{S_j} = \sum_i S_i \left( \frac{\tau_{i,j}}{\sum_j \tau_{i,j}} \right) / N \qquad (18)$$

where $N$ is the number of samples. Compared with CPF and NPR, TSA considers the entire air mass history rather than just the wind direction at the receptor.

*Data Needs*

TSA requires the time series of pollutant concentration or source contribution at the receptor site, and back trajectories initiated over the site during the sampling duration. Trajectory is usually calculated once every hour so TSA is more suitable for analyzing measurements of $>$ 1-hr resolution.

*Output*

TSA reports the average pollutant concentration or source contribution as a function of wind direction based on back trajectory calculations.

*Strengths*
- Infer the direction of sources or factors relative to the sampling site.
- Provide verification for the source identification made by factor analysis method.
- Account for air mass transport over hundreds to thousands of kilometers and on the order of several days.
- Can represent plume spread from vertical wind shear at different hours of day by adjusting the initial height of back trajectories.

*Weaknesses*
- Need to generate and analyze the back trajectory data.
- Uncertainty in back trajectory calculation increases with its length in time.
- Source contribution depends on not only trajectory residence time but also entrainment efficiency, dispersion, and deposition.
- Difficult to resolve the direction of more localized sources.

**PSCF[140]**

PSCF requires the calculation of air parcel back trajectory, which is often accomplished using the HY-SPLIT model.[115,139] HY-SPLIT version 4.5 is available at *http://www.arl.noaa.gov/ready/hysplit4.html*.

*Principle*

Ensemble air parcel trajectory analysis refers to the statistical analysis on a group of trajectories to retrieve useful patterns regarding the spatial distribution of sources. Uncertainties associated with individual trajectory calculations largely cancel out for a sufficient number of trajectories or trajectory segments. As a popular ensemble back trajectory analysis, PSCF estimates the probability that an upwind area contributes to high pollutant concentration or source contribution. Back trajectories are first calculated for each sample at the receptor site. To determine the PSCF, a study domain containing the receptor site is divided into an array of grid cells. Trajectory residence time (the time it spends) in each grid cell is calculated for all back trajectories and for a subset of trajectories corresponding to "high" pollutant concentration or source contribution at the site. PSCF in cell ($i,j$) is then defined as:

$$PSCF_{i,j} = \frac{\text{Sum of "high" residence time in cell } (i, j)}{\text{Sum of all residence time in cell } (i, j)} \qquad (19)$$

The criterion for high pollutant concentration or source contribution is critical for the PSCF calculation. The 75th or 90th percentile of the concentration or factor is often used.[113,141,142] Residence time can be represented by the number of trajectory end points in a cell.

*Data Needs*

Similar to TSA, PSCF calculation requires the time series of pollutant concentration or source contribution at the receptor site, and back trajectories initiated over the site during the sampling period. Trajectories should be calculated with 1- to 3-h segment to reduce the uncertainty from interpolation (if needed).

*Output*

PSCF reports the probability that an upwind area contributes to high pollutant concentrations or source contribution at the downwind receptor site. The results are often presented as a contour plot on the map. A high probability usually suggests potential source region (*e.g.*, Figure 4b).

*Strengths*
- Infer the location of sources or factors relative to the sampling site.
- Provide verification for the source identification made by factor analysis method
- Account for air mass transport over hundreds to thousands of kilometers and on the order of several days.
- Resolve the spatial distribution of source strength (qualitatively).

*Weaknesses*
- Need to generate and analyze the back trajectory data.
- Need to correct for the central tendency (residence time always increases toward the receptor site regardless of source contribution).
- Uncertainty in back trajectory calculation increases with its length in time.
- Source contribution depends on not only trajectory residence time but also entrainment efficiency, dispersion, and deposition.
- Difficult to resolve the location of more localized sources.

**Table 1d. (cont.)**

| Meteorology-Based Method | | |
|---|---|---|
| **Receptor Model** | **Description** | **Strengths and Weaknesses** |
| SQTBA[117,143] SQTBA requires the calculation of air parcel back trajectory, which is often accomplished using the HY-SPLIT model.[115,139] HY-SPLIT version 4.5 is available at *http://www.arl.noaa.gov/ready/hysplit4.html.* | *Principle* SQTBA is another type of ensemble air parcel trajectory analysis. The concept of SQTBA is to estimate the "transport field" for each trajectory ignoring the effects of chemical reactions and deposition. Back trajectories are first calculated for each sample at the receptor site, and a study domain containing the receptor site is divided into an array of grid cells. SQTBA assumes that the transition probability that an air parcel at $(x', y', t')$, where $x'$ and $y'$ are spatial coordinates and $t'$ means time, will reach a receptor site at $(x, y, t)$ is approximately normally distributed along the trajectory with a standard deviation that increases linearly with time upwind[144,145], thus: $$Q(x, y, t \mid x', y', z') = \frac{1}{2\pi(at')^2} \exp\left[ -\frac{1}{2}\left( \left(\frac{X - x'(t')}{at'}\right)^2 + \left(\frac{Y - y'(t')}{at'}\right)^2 \right) \right]$$ (20) where $(X,Y)$ is the coordinate of the grid center, $a$ is the dispersion speed, and $x'(t')$ and $x'(t')$ represent the trajectory. The probability field, $Q$, for a given trajectory is then integrated over the upwind period, $\tau$, to produce a two-dimensional "natural" (nonweighted) transport field: $$T_k(x, y \mid x', y') = \frac{\int_{-\tau}^{0} Q(x, y, t \mid x', y', z')}{\int_{-\tau}^{0} dt'}$$ (21) After the transport field for each trajectory is established, they are weighted by the corresponding pollutant concentration or source contribution at the receptor site and summed to yield the overall SQTBA field.[117] *Data Needs* SQTBA requires the time series of pollutant concentration or source contribution at the receptor site, and back trajectories initiated over the site during the sampling period. Trajectories should be calculated with 1- to 3-hr segment to reduce the uncertainty from interpolation (if needed). *Output* SQTBA put more weight on trajectories associated higher pollutant concentration or source contribution and therefore the resulting field may imply the major transport path. | *Strengths* ● Imply the location of sources or factors relative to the sampling site. ● Account for air mass transport over hundreds to thousands of kilometers and on the order of several days. ● Resolve the spatial distribution of source strength (qualitatively). *Weaknesses* ● Need to generate and analyze the back trajectory data. ● Need to correct for the central tendency (residence time always increases toward the receptor site regardless of source contribution). ● Need to estimate dispersion velocity. ● Involve complicated calculations. ● Physical meaning of the SQTBA field is unclear. ● Difficult to resolve the location of more localized sources. |
| RTWC[146] RTWC requires the calculation of air parcel back trajectory, which is often accomplished using the HY-SPLIT model.[115,139] HY-SPLIT version 4.5 is available at *http://www.arl.noaa.gov/ready/hysplit4.html* | *Principle* As an ensemble air parcel trajectory analysis, RTWC requires back trajectories calculated for each sample at the receptor site, and a study domain containing the receptor site divided into an array of grid cells. RTWC assumes that no major pollutant sources are located along "clean" (associated with low pollutant concentrations) trajectories and that "polluted" trajectories picked up emissions along their paths. In practice, RTWC distributes pollutant concentrations at the receptor to upwind grid cells along the back trajectories according to the trajectory residence times in those cells[117,146]: $$S_{i,k} = S_k \frac{\text{resident time in cell } i}{\text{average residence time in each cell}}$$ (22) where $S_k$ is the pollutant concentration or source contribution determined upon the arrival of trajectory $k$ and $S_{i,k}$ is the redistributed pollutant concentration or source contribution for cell $i$ upwind. RTWC is known for the problem of "tailing effect," i.e., spurious source areas can be identified when cells are crossed by a very small number of trajectories. Although some corrections were proposed[147] these approaches are purely empirical. | *Strengths* ● Imply the location of sources or factors relative to the sampling site. ● Account for air mass transport over hundreds to thousands of kilometers and on the order of several days. ● Resolve the spatial distribution of source strength (qualitatively). *Weaknesses* ● Need to generate and analyze the back trajectory data. ● Need to correct for the central tendency and tailing effect. ● The amount of emission entrainment should not be proportional to the residence time of trajectories (so there is no linear relationship between RTWC field and source strength). ● Physical meaning of the RTWC field is unclear. ● Difficult to resolve the location of more localized sources. |
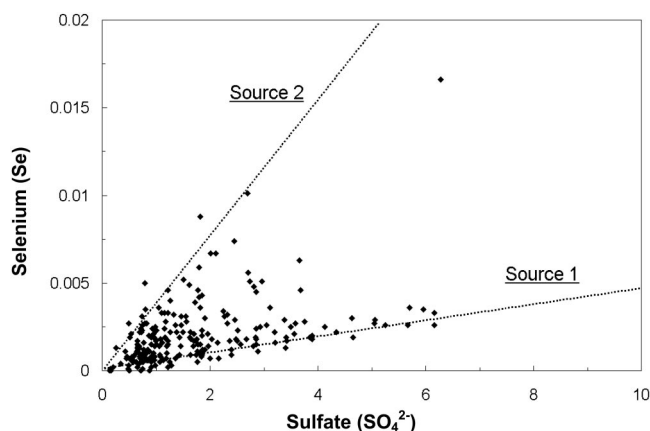
**Figure 1.** Scatter plot of concentrations of two species. These are real concentrations measured at Fort Meade, MD, between July 1999 and July 7, 2001.[118] The *x*-axis is 24-hr $SO_4^{2-}$ and *y*-axis is 24-hr Se. The unit is μg/m³. The clear edges indicate two sources with different $SO_4^{2-}$/Se ratio.

$C^{13}/C^{14}$ for biogenic emissions,[49] and selenium (Se) for coal-fired power plant emissions.[50] Water-soluble potassium ($K^+$) and levoglucosan are markers for vegetative burning, and cholesterol is a marker for meat cooking. Silicon (Si) and calcium (Ca) enrichments often differ between local soil and global crustal materials. The number of sources can be inferred from scatter plots of two or more species (e.g., Figure 1). Three variants of marker or tracer solutions are described in Table 1c, namely enrichment factor (EF), non-negative least squares (NNLS), and fractional aerosol coefficient (FAC). Multiple linear regression (MLR) is also used for tracer solutions. FPEAK and $F_{key}$ are codes in the PMF program, being used as examples of such codes in this paper.

The component *n* in eq 1 allows source contribution estimates to be classified by transport direction, thereby allowing a further limitation on the possible contributors. For daily or longer sample durations, this has not been very accurate because the transport patterns usually change substantially within the day. However, several Supersite examples show relationships of receptor concentrations to long-range transport occurring over several days. Transport classification becomes more reasonable with shorter sample durations such as those achieved from several of the continuous speciation methods. The conditional probability function (CPF) and nonparametric regression (NPR) use only local wind direction and wind speed data measured near the surface, whereas trajectory sector analysis (TSA), potential source contribution function (PSCF), simplified quantitative transport bias analysis (SQTBA), and residence time-weighted concentrations (RTWC) use air parcel back trajectories calculated often by the hybrid single-particle Lagrangian integrated trajectories (HY-SPLIT) model for large-scale transport (see descriptions in Table 1d).

## RECEPTOR MODELING PROCEDURES
Reff et al.[51] suggest three steps: (1) preparing data to be modeled, (2) processing the data to develop a feasible and robust solution, and (3) interpreting the results. These are included in the eight-step procedure recommended by Watson et al.,[37] providing a framework for most Supersite source apportionment studies:

### Step 1: Form a Conceptual Model
As the first step, receptor modelers must establish conceptual models[52] that provide qualitative descriptions of the number and types of sources that might affect a receptor. The conceptual model should guide the siting of measurement systems, the variables to be measured, and the sampling periods and durations. Unfortunately, receptor models are usually applied to data available from networks that were designed for other purposes, but they have always added value to these measurements.

### Step 2: Compile Emissions from Emission Inventories
Receptor models need to be supplied with sources that are potential contributors. A receptor model inventory requires only source types, not the locations and emission rates of specific sources, although the locations and emission rates of point sources help interpret the receptor model results. Source profiles should be closely associated with the specific emission sources, and generic source profiles acquired from other locations should be used only as a last resort. The lack of representative source profiles for a large number of chemical species and cost-efficient means of acquiring these profiles was not well addressed during the Supersites Program, but some progress was made.

### Step 3: Characterize Sources and Source Profiles
More than 1000 measured source profiles for geological, motor vehicle, industrial, vegetative burning, and cooking emissions for different PM size fractions were compiled and made available as part of EPA SPECIATE Version 4 database (www.epa.gov/ttn/chief/software/speciate/).[10] More specific source markers need to be accurately and efficiently measured at source and receptor to improve SCEs for all of the models. Thermally derived C fractions,[53–56] which are a free byproduct of thermal/optical C analyses, have been found useful for distinguishing between diesel exhaust and other C source types.[57–61] For example, high-temperature EC (e.g., EC2, EC evolved at 740 °C in a 98% helium [He]/2% oxygen [$O_2$] atmosphere) dominates particulate emissions from diesel engines.[54] Specific organic markers (Table 2) have also proven useful in PMF modeling studies,[63–66] but advances are needed to make them more available in source and receptor samples.[61]

Modern PM source profiles that include organic compounds were obtained and applied for source apportionment at Fresno[64] and Pittsburgh.[67–69] Figure 2 illustrates some of these source profiles that contain at least 35 common species (elements, ions, and thermal C fractions) as well as 25 organic markers. How the markers are used to distinguish various source emissions is tabulated in Tables 2 and 3.

### Step 4: Create a Material Balance for Each Sample and Site
Reconstructed mass is a weighted sum of species (i.e., sulfate [$SO_4^{2-}$] nitrate [$NO_3^-$], ammonium [$NH_4^+$; where available], geological material [with elemental multipliers

**Table 2.** Summary of organic marker species for major PM$_{2.5}$ sources.[64,74]

| Organic Aerosol Sources | Highly Specific Organic Markers | Important Organic Markers |
|---|---|---|
| Natural gas combustion | | PAHs |
| Gasoline-powered engine exhaust | | Hopanes and steranes |
| | | PAHs |
| | | EC |
| Diesel engine exhaust | | Hopanes and steranes |
| | | EC |
| Fuel oil combustion | | Hopanes and steranes |
| | | EC |
| | | Ni |
| | | V |
| Coal combustion–uncontrolled | Picene | Hopanes and steranes |
| | | PAHs |
| | | EC |
| Biomass burning | Levoglucosan | Resin acids |
| | | Methoxyphenols |
| Meat cooking | Cholesterol | Tetradecanoic acid |
| | | Pentadecanoic acid |
| | | Hexadecanoic acid |
| | | Heptadecanoic acid |
| | | Octadecanoic acid |
| | | Palmitin |
| | | Palmitoleic acid |
| | | Oleic acid |
| | | Stearin |
| Vegetative detritus | | *n*-Nonacosane |
| | | *n*-Triacontane |
| | | *n*-Hentriacontane |
| | | *n*-Dotriacontane |
| | | *n*-Tritriacontane |
| SOA | 1,2-Benzenedicarboxylic acid | Not well defined |
| | Methyl-1,2-benzenedicarboxylic acid | |
| Cigarette smoke | Iso-nonacosane | |
| | Anteiso-triacontane | |
| | Iso-hentriacontane | |
| | Anteiso-dotriacontane | |
| | Iso-tritriacontane | |

*Notes:* PAH = polycyclic aromatic hydrocarbon; SOA = secondary organic aerosol.

to account for unmeasured oxides]), EC, and organic carbon (OC; weighted by a multiplier to account for unmeasured hydrogen, oxygen, and other elements). The reconstructed mass is compared with gravimetric mass often measured on Teflon-membrane filters. Excesses or deficits in mass indicate unidentified species and/or potential measurement errors. Mass closure should be part of a data validation procedure that also includes cation/anion balance and comparison between different measurement techniques for duplicate species (e.g., sulfur [S] and chlorine [Cl] by X-ray fluorescence [XRF] vs. SO$_4^{2-}$ and chloride [Cl$^-$] by ion chromatography; potassium [K] and sodium [Na] by X-ray fluorescence vs. water-soluble K$^+$ and sodium [Na$^+$] ions by ion chromatography or atomic absorption spectroscopy).[70]

Data validation allows receptor modelers to identify outliers that should be excluded from model inputs.[44,71] Uncertainty estimates may need to be adjusted, because analytical uncertainties may not reflect the true uncertainty. As described in Table 1, a and b, the effective variance and PMF solutions use uncertainties as input data. The PMF solution needs more adjustment because it does not use source profile uncertainties, which usually dominate the weighting in the effective variance solution. Measurement precision (i.e., based on analytical minimum detectable limit [MDL] and replicate analysis) alone does not account for source profile variability. A poor mass closure for lightly loaded samples may result from insufficient adjustments for organic sampling artifacts.[71–73] Because back-up filters are not always available allowing for an accurate correction and error propagation for the sampling artifact, a larger uncertainty may be assigned to OC, total carbon (TC), and PM mass measurements.

## Step 5: Descriptive Data Analysis
Descriptive data analysis summarizes the data (i.e., maximum, minimum, mean, median, 98th percentile by site, year, and/or season) and may also include time series plots, temporal correlations, spatial correlations, and principal component analysis.[74] Data analysis should identify samples of special interest (e.g., air quality standard exceedances), similarities and differences among measurement locations, and potential modifications to

**Figure 2.** Source profiles of: (a) paved road dust (PVRD), (b) gasoline vehicle exhaust (GAS), (c) diesel vehicle exhaust (DIESEL), (d) cooking (COOK), and (e) RWC (BURNING) used for the Fresno Supersite PM$_{2.5}$ apportionment.[64] Uncertainties (UNC) of the profiles are indicated by triangles. Indeno[123]:Indeno(1,2,3-cd)pyrene; Benzo(g,h,i) p: Benzo(g,h,i) perylene; Ster35: C27-20S-13β(H),17a(H)-diasterane; ster45: C27-20R5a(H),14a(H),17a(H)-cholestane and C29-20S13β(H),17a(H)-diasterane; ster48: C28-20S5a(H),14β(H),17β(H)-ergostane; ster49: C28-20R5a(H),14a(H),17a(H)-ergostane; hop17: 17a(H),21β(H)-30-Norhopane; hop19: 17a(H),21β(H)-Hopane; hop24: 22S-17a(H),21β(H)-30,31-Bishomohopane; hop26: 22S-17a(H),21β(H)-30,31,32-Trisomohopane.

the conceptual model. High correlation coefficients usually indicate origin from a common source or effects of common meteorology on several sources. High spatial correlations among chemical components indicate urban or regional source influences, whereas low correlations indicate influences from more local emitters.

**Step 6: Calculate Source/Factor Contributions**
Input data should be selected, documented, and formatted for model input in such a way that they are available for others to understand and use them for verification. Many of the reported receptor model results are not reproducible by others, especially for UNMIX and PMF solutions, because the datasets are not sufficiently documented. Reff et al.[51] point out inconsistencies in the literature regarding the preparation of data and uncertainty matrices for PMF. These include the treatment of duplicate measurements and values that are missing or below MDL. Several equations that are functions of concentrations, analytical precisions, and MDLs have been

used to create the uncertainty matrix. Kim and Hopke[75] point out that analytical precisions are estimated differently by different laboratories, thereby influencing the PMF results. The variability in source profiles (i.e. $\sigma_{Fijm}$), is not considered explicitly in UNMIX and PMF solutions, although source factor uncertainties can be assigned by bootstrapping methods.[76,77]

UNMIX, PMF, and effective variance CMB software offer several solution options (e.g., robust vs. nonrobust), species included, number of factors or profiles, and software implementation[78] that result in different results from the same data. Receptor model solutions can be unstable (i.e., multiple or no solutions) when two or more source profiles are similar, or when ambient concentrations are highly correlated in time and space owing to similar meteorological transport and dispersion.[79,80] The assignment of a source factor to a source type often represents a stretch of the imagination when compared with measured source profiles. Therefore, sensitivity tests that

**Table 3.** Summary of inorganic marker species for major PM$_{2.5}$ sources.[148]

| Source | Descriptions | Marker Species |
|---|---|---|
| Road dust | Paved and unpaved road dust by traffic | Al, Si, K, Ca, Fe |
| Fugitive windblown dust | Farm lands; pasture lands; unpaved roads | Al, Si, K, Ca, Fe |
| Construction and demolition | Building; road construction dust | Al, Si, K, Ca, Fe |
| Farming operations | Tilling, harvesting, and growing; livestock | OC, NH$_3$, NH$_4^+$ |
| Mobile | On-road gasoline and diesel vehicles; off-road mobile | OC, EC, EC1[a], EC2[a], NH$_3$, S, Fe, Zn |
| Cooking | Indoor and outdoor cooking | OC, EC |
| Fires | Wild forest fires | OC, EC, K$^+$, Cl$^-$ |
| Waste burning and disposal | Agriculture burning; incineration; prescribed burning | OC, EC, K$^+$, As, Pb, Zn |
| Industrial fuel combustion | Electric utility, cogeneration, oil and gas production; petroleum refining; manufacturing and industrial; food and agriculture processing; service and commercial | SO$_4^{2-}$, Se, V, Ni, OC, EC |
| Residential fuel combustion | Wood combustion; cooking and space heating | OC, EC, K$^+$, Cl$^-$ |
| Industrial processing | Chemical; food and agriculture; mineral processes; metal processes; wood and paper; glass and related products; electronics | Zn, Pb, Cu, Mn, As, Hg |
| Sea salt | Marine aerosol | Na$^+$, Cl$^-$ |
| NH$_4$NO$_3$ | Secondary aerosol | NO$_3^-$, NH$_4^+$ |
| [NH$_4$]$_2$SO$_4$ | Secondary aerosol | SO$_4^{2-}$, NH$_4^+$ |
| SOA | Secondary aerosol | OC |

*Notes:* [a]EC1 and EC2 are the EC fractions evolved at 740 and 840 °C in a 98% helium/2% oxygen atmosphere, following the IMPROVE_A protocol.[54]

assess variability of the results because of different combinations of sources and/or species in the model are important. Performance measures, such as variance explained (r$^2$) and residual ($\chi^2$) of measured-versus-calculated concentrations help validate model outputs.[81] Many different calculations need to be made before "the best" results are reported in reports and publications.

### Step 7: Evaluate Limiting Precursors of Secondary Aerosol

Source characteristics may change during transport to a receptor, the most common being changes of sulfur dioxide (SO$_2$) and oxides of nitrogen (NO$_x$) gases to SO$_4^{2-}$ and NO$_3^-$ particles as well as the Cl$^-$ depletion of the sea-salt aerosol. These changes can be simulated with aerosol evolution models.[82] Secondary ammonium sulfate ([NH$_4$]$_2$SO$_4$) and ammonium nitrate (NH$_4$NO$_3$) involve ammonia (NH$_3$) from noncombustion sources that may be a limiting precursor. Reductions in SO$_4^{2-}$ might increase NH$_4$NO$_3$ concentrations by freeing up NH$_3$.[83] Chemical equilibrium receptor models determine the extent to which one precursor needs to be diminished to achieve reductions in NH$_4$NO$_3$ levels. The potential effects of planned sulfur reductions can be addressed from Supersite measurements.[84] Using an observational (i.e., receptor) box model, Vayenas et al.[85] estimated that a 50% reduction of SO$_4^{2-}$ in winter would lead

to only an 8% reduction in PM$_{2.5}$ mass concentration in Pittsburgh. For a 50% reduction in NH$_3$ availability, PM$_{2.5}$ would be reduced by 29%.

### Step 8: Reconcile Source Contributions with the Conceptual Model and Report Uncertainties in the Source Contribution Estimates

A receptor analysis can be considered valid when: (1) the receptor model is determined to be applicable; (2) the performance measures are within target ranges; (3) there are no large deviations from model assumptions; (4) sensitivity tests reveal acceptable instability or consistency problems; and (5) there is consistency with the conceptual model, or reasonable modifications to the conceptual model are revealed.[38] Receptor modelers need to resist the temptation to propose implausible reasons for their results that have not been verified by other means: doubting the source contribution estimates rather than picking the best ones, and using discrepancies between different sets of input data and model parameters to estimate uncertainties.

An example of sensitivity testing for PM$_{2.5}$ source apportionment at Fresno[64] is given in Table 4. Initial tests with different combinations of source profiles were done to determine which profiles best explain the data. This test reveals that including organic markers in the single sample CMB solution improves the distinction between gasoline and diesel vehicle emissions and

**Table 4.** Sensitivity test for source apportionment at the Fresno Supersite.

| | Source Contribution Estimates (μg/m$^3$) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | PVRD | GAS | DIES | BURN-H | BURN-S | COOK | MARINE | AMSUL | AMNIT | PMASS | r$^2$ | $\chi$ |
| 1[a] | 0 | 1.9 ± 1.3 | 4.4 ± 1.5 | 16 ± 3 | 5.8 ± 6.2 | 20 ± 5 | 0 | 1.1 ± 0.4 | 18 ± 2 | 89 | 0.96 | 0.6 |
| 2[b] | 0 | 0 | 4.8 ± 1.5 | 15 ± 3 | 7.0 ± 6.4 | 23 ± 6 | 0 | 1.3 ± 0.3 | 18 ± 2 | 90 | 0.98 | 0.7 |

*Notes:* These trial runs apply to average concentrations for the 12:00 a.m. to 5:00 a.m. Pacific Standard Time samples during the winter intensive period.[65] PVRD = paved road dust (PVRD); GAS = gasoline vehicle exhaust; DIES = diesel vehicle exhaust; BURN-H = hardwood burning; BURN-S = softwood burning; COOK = cooking; AMSUL = secondary ammonium sulfate; AMNIT = secondary ammonium nitrate. r$^2$ and $\chi$ are performance measures. [a]With organic markers; [b]without organic markers.

allows a more precise estimate of the cooking source contribution. However, organic markers were not required to precisely estimate the RWC contribution and did not increase the precision of the softwood burning contribution, even though there were differences among the hardwood and softwood RWC markers such as levoglucosan, 4-allyl-guaiacol, and syringaldehyde.

As noted before, personal judgment is used to link UNMIX and PMF source factors with measured source profiles, and the rationale for these judgments is not always apparent. It usually relies on the presence or absence of reported organic and inorganic source markers such as those identified in Tables 2 and 3. However, some of the source factors contain multiple markers that indicate a possible mixing of several source types. Alternatively, there is a factor with a single marker (i.e., the famous "zinc [Zn]-rich" factor) that can only remotely be related to a single source (the Zn-rich factor is often attributed to tire wear or incineration, even though one would expect other markers to come from these source types). UNMIX and PMF source factors need to be accompanied by a plausibility analysis, in which the source factors are compared with measured source profiles before they are assigned to a source type.

When the $T_{ijklmn}$ in eq 1 is assumed to equal one, or a constant for all transport directions, actual variability of the source profiles in space and time can still influence the outcome. This usually leads to unacceptable $r^2$ and/or $\chi^2$ and unrealistic source apportionment results.[86] In those cases, receptor model inputs can be confined within a shorter period and/or a smaller area. Chen et al.[45] observed that seasonal variability for $PM_{2.5}$ source profiles was larger than spatial variability around Fresno. They used speciated $PM_{2.5}$ data from 23 sites in California's central valley for winter and non-winter period UNMIX and PMF solutions. They found that derived source factors were more similar to measured profiles than would be obtained from year-long samples at a single location. This is consistent with Cadle et al.,[87] who found different motor vehicle source profiles (in terms of OC/EC ratios) between summer and winter, attributable to larger contributions of cold starts during winter. Chen et al.[45] evaluated source apportionment results based on: (1) site-specific residuals between the measured and calculated concentrations; 2) comparability of motor vehicle and RWC factors against source profiles obtained from emission tests; (3) spatiotemporal variations of the factors' strengths; and (4) edges in bi-plots of key marker species. Clear edges in the OC versus EC scatter plot (Figure 3) correspond to the PMF motor vehicle, RWC, and cooking factors.

## HOW WELL CAN WE IDENTIFY AND QUANTIFY SOURCE CONTRIBUTIONS USING RECEPTOR MODELS?

Source apportionment studies carried out as part of the Supersites Program are summarized in Table S-1 published online at http://secure.awma.org/journal/pdfs/2008/2/10.3155-1047-3289.58.2.265_supplmaterial.pdf. It was intended that these studies would take advantage of the comprehensive Supersite measurements at various time resolutions[88,89] and attempt to use appropriate modeling
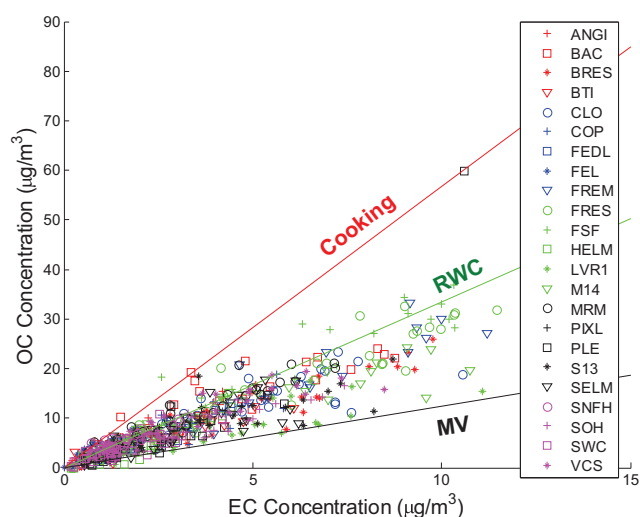


**Figure 3.** $PM_{2.5}$ OC vs. EC concentration by site (a total of 23 sites) in California's San Joaquin Valley during the high_$PM_{2.5}$ period (November–January). Data were collected as part of the California Regional $PM_{10}$/$PM_{2.5}$ Air Quality Study (CRPAQS).[74] Solid lines indicate the OC/EC ratios for the motor vehicle (MV), RWC, and cooking factors determined by PMF analysis.[45] Clear edges suggest that source profiles are consistent among different sites.

techniques. SCEs were made for several PM size fractions and for VOCs. UNMIX and PMF solutions were applied, with the reason given that source profiles were not available. Several of the studies applied more than one of the UNMIX, PMF, and CMB solutions to the same or similar datasets and used the similarities and differences to evaluate the accuracy and precision of the SCEs. Table 5 compares results from different studies.

PM at the eastern Supersites (e.g., Baltimore, New York, Pittsburgh, and St. Louis) was dominated by secondary $(NH_4)_2SO_4$, mostly deriving from the Midwest. Fugitive dust contributions to $PM_{2.5}$ were small, and secondary $NH_4NO_3$ contributions were variable, with higher values in winter. The highest C concentrations were often found when there was influence from wildfires, such as those in Canada.[90] SOA was determined to be an important contributor on the basis of the enrichment of OC over the presumed EC contribution from primary sources.

### Intermodel Comparability

$PM_{2.5}$ source apportionment was conducted for the Pittsburgh Supersite with UNMIX,[65] PMF,[65,91] and ME.[46] These solutions used different measurements as input for the July 2001 to July 2002 study period. UNMIX used fewer species (12) than PMF (22) because PMF solutions could give less weight to outliers that biased the UNMIX solution. Pekney et al.[65] reported six common factors: (1) crustal material (Ca, titanium [Ti], iron [Fe]); (2) $NO_3^-$; (3) $SO_4^{2-}$; (4) manganese (Mn)/Fe/Zn; (5) specialty steel (chromium [Cr], molybdenum [Mo]); and (6) cadmium [Cd]. An additional four factors, represented by gallium (Ga), lead (Pb), Se, and primary OC/EC, were identified by PMF. The UNMIX factors explained nearly 100% of $PM_{2.5}$ mass whereas PMF factors accounted for 78% of $PM_{2.5}$. The mobile source category was not found by the UNMIX solution, even though vehicle exhaust probably contributes at any urban site. The primary OC/EC factor from

**Table 5.** Comparison of source/factor contributions to PM at Supersites.

| Locations | Model | Marine | Crustal Material | Motor Vehicle[a] | Gasoline Vehicle | Diesel Vehicle | Biomass Burning | Cooking | Secondary Sulfate | Secondary Nitrate | MISC 1 | MISC 2 | MISC 3 | PM2.5 (μg/m³) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pittsburgh[65] | UNMIX | | 13 | | | | | | 68 | 14 | 4[f] | 1[g] | 1[h] | |
| Pittsburgh[65] | PMF | | (13) | 11 | | | (3) | | (68) | (14) | (4[f]) | 2[i] | 2[j] | |
| Pittsburgh[46] | ME | | k | k | | | k | | k | k | k,l | g,k | | |
| Pittsburgh[48] (OC) | PT(EC) | | | | | | | | | | 25[e] | | | |
| Fresno (23-site average in the SJV)—winter[45] | UNMIX | 0 | 3 | 15 | | | 24 | 5 | | 51 | 2[b] | | | 42.2 |
| Fresno (23-site average in the SJV)—winter[45] | PMF | 0 | 5 | 10 | | | 23 | 3 | | 48 | 2[b] | 9[c] | | 42.2 |
| Fresno winter[101] | CMB | 0.1 | 0.1 | | 8 | 6 | 31 | 5 | 2 | 32 | | | | 72 |
| Fresno winter episode I[101] | CMB | | 0 | | 3.8 | 12.7 | 63.1 | 13 | 3.4 | 27.7 | 1.4[d] | 3.1[e] | | 76.4 |
| Fresno winter episode II[45] | CMB | | 1.8 | | 1.9 | 9.3 | 26.6 | 2.7 | 7 | 44.1 | 7.5[e] | | | 55.3 |
| Fresno (23-site average in the SJV)—non-winter[45] | UNMIX | 10 | 16 | 25 | | | | 5 | | 38 | 5[b] | | | 8.4 |
| Fresno (23-site average in the SJV)—non-winter[45] | PMF | 7 | 19 | 13 | | | | 9 | 7 | 36 | 4[b] | 6[c] | | 8.4 |
| Houston[63] | CMB | | 10.6 | | 12.8 | 17.3 | 1.5 | 5.9 | | | 2.1[m] | 6.7[n] | | 21.6 |
| Houston—midday[149] | CMB | | 4.8 | | 17.8 | 64.9 | | | | | | | | 99 |
| Houston—afternoon[149] | CMB | | 5.5 | | 28 | 65.3 | | | | | | | | 103 |
| Houston[106] (OC) | PT (EC) | | | | | | | | | | 5–10[e] | | – | |
| Baltimore—summer episode[96] | PMF | | 3.5 | | 8.1 | 1.9 | 30.9 | | 14.7 | 18.6 | 10.3[j] | 2.9[n] | 7.9[o] | |
| Baltimore[118] | UNMIX | | | 15 | | | 13 | | 45 | 17 | 8[j] | 2[p] | | 13 |
| New York[150] | PMF | 14.6 | 35 | 4.4 | | | | | 64.6 | 10.8 | | 9.1[n] | | 11.8 |
| New York—rural[151] | PMF | 14.6 | | | | | 8.8 | | 56 | 9.2 | 2.0[q] | 2.0[c] | 4.2[r] | |
| St. Louis[152] | PMF | | 4.2 | | 16.4 | 2.1 | | | 52.2 | 15.3 | 0.5[q] | 1.3[c] | 6.8[g] | 16.1 |
| St. Louis—STN site A[153] | PMF | | 15.4 | 17.2 | | | | | 39.9 | 19.7 | 3.4[q] | 3.0[c] | 0.3[g] | 16.4 |
| St. Louis—STN site B[153] | PMF | | 3.2 | | 21.1 | 5 | 2.4 | | 36.1 | 12.9 | 11.6[s] | 2.6[p] | 3.2[j] | 15.5 |
| Atlanta[154] | PMF | | 3.1 | | 5.3 | 11.6 | 22 | | 29.4 | 9.2 | 1.6[j] | 6.9[p] | | 16.7 |

*Notes:* Contributions are presented as percentage (%) of PM2.5 mass concentration (in μg/m³). Numbers in parentheses are estimated values. A question mark indicates unreported values. PT = primary tracer method; MSC1–3 = miscellaneous sources. [a]Combined gasoline and diesel vehicle exhaust; [b]dairy; [c]zinc; [d]natural gas combustion; [e]secondary organics; [f]Fe, Mn, Zn; [g]specialty steel; [h]CD; [i]Pb; [j]Se (coal burning); [k]unknown or undetermined contribution; [l]coke plant; [m]vegetative detritus; [n]oil combustion; [o]incinerator; [p]industry; [q]copper smelter; [r]nickel smelter; and [s]Ca.

PMF accounted for 88 and 9% of ambient OC and EC, respectively, which was assigned to the motor vehicle source type. Most of the OC and EC was distributed among the other factors in the UNMIX solution.

Comparisons of UNMIX and PMF SCEs at the Fresno Supersite during the California Regional PM10/PM2.5 Air Quality Study (CRPAQS)[45] also found that fewer input species could be used for a stable UNMIX solution. Nevertheless, the UNMIX and PMF source factors, their association with measured source profiles, and SCEs agreed. Compared with Pittsburgh, much more PM2.5 in the central California valley came from secondary NH4NO3, vegetative burning, and farming activities. The mobile source factor was evident in both the UNMIX and PMF solutions.

The contrast between the Fresno and Pittsburgh results show the danger of extrapolating agreement between different solutions from one situation to others. The application of UNMIX and PMF solutions in Fresno gives more confidence to both sets of results, which are also consistent with the organic marker CMB solution at Fresno during winter.[64] The discrepancies found in Pittsburgh are just as valuable, because they indicate a larger uncertainty in the SCEs that needs to be considered by decision-makers.

Pekney et al.[65] enhanced their analysis by including organic markers in the PMF model. A few organic classes have been shown to be effective source indicators, including hopanes for vehicle exhaust, levoglucosan for vegetative burning, and *n*-alkanes for vegetative detritus (Table 2). Because organic markers were not measured in all samples at Pittsburgh (e.g., 24% of the sampling days), gaps were filled with the geometric mean concentration and assigned an uncertainty of four times the average precision, as suggested by Polissar et al.[92,93] Although the effect of this approximation has not been systematically evaluated, Pekney et al.[65] found that all of the hopanes were correlated with a primary OC/EC factor, and this was assigned to mobile source emissions. The final PMF solution contained 11 factors, with the primary OC/EC factor split into two factors: one associated with vehicle exhaust and road dust, and the other associated with vegetative burning, cooking, and vegetative detritus. These two factors usually cannot be separated based only on elemental, ionic, and thermal C fraction measurements.

Robinson et al.[67–69] demonstrated that the ratio of marker species in a source profile, when compared with those from the same and/or different source types and from ambient samples, help interpret the source variability and suggest the most important sources in a region. The comparison of the ratio of Benzo(g,h,i)perylene/Benzo(e)pyrene against the ratio of Indeno(1,2,3,-cd)-pyrene/Benzo(e)pyrene (in a ratio-ratio plot) at Pittsburgh most resembled ratios from coke (coal product) production, suggesting influence from nearby coke plants. Contributions from up to three distinct sources and the atmospheric aging of single source profiles was implied from ratio-ratio plots, thus providing some indication of $T_{ijklmn}$

in eq 1. These sources are not necessarily major contributors to $PM_{2.5}$ or OC mass, but they may be important hazardous air pollutant contributors.

Source apportionment using only organic markers did not quantify vegetative burning contributions in Pittsburgh. There were no strong correlations among three organic markers for vegetative burning: levoglucosan, resin acids, and syringhaldehyde at Pittsburgh.[68] Robinson et al.[68] attribute this to the poorly controlled, highly variable nature of biomass combustion. They estimated that vegetative burning contributed approximately 10% of OC in winter. Chow et al.[64] observed that organic markers were not critical for resolving biomass burning contributions in Fresno as long as $K^+$ was included in the effective variance solution.

Cooking contributions were reported in a few studies using organic markers.[64,69,94] Robinson et al.[69] identified three uncertainty areas for cooking contributions at Pittsburgh: (1) the markers (e.g., palmitoleic acid and cholesterol) were highly uncertain and variable in the reported source profiles; (2) the ambient oleic acid/palmitoleic acid ratio was more than a factor of 10 greater than all published source profiles in Pittsburgh; and (3) three source profile combinations yielded equally good performance statistics, but cooking contributions varied by a factor of 9 (3–27% of OC). Chow et al.[64] estimated cooking contributions of 5–19% of $PM_{2.5}$ at Fresno during winter episodes with very large propagated uncertainties.

Organic markers appear to be useful for separating vegetative burning and fossil fuel combustion using PMF. They might also further distinguish gasoline, diesel, and cooking contributions, as suggested by Chow et al.[64]

Zhou et al.[46] analyzed high-time resolution measurements with ME at Pittsburgh during July 2001. As the time resolution increases, sufficient data for the analysis can be acquired in a relatively short time period during which variations in source characteristics are minimal. However, the study period also needs to be long enough to capture variability in source contributions. ME2, rather than PMF, was used to analyze data from instruments with different sampling durations. Estimated source contributions were averaged to a consistent measurement interval. This can only be achieved under the framework of ME2. The analysis used 24 inorganic species and identified six sources, including motor vehicles, geological dust, steel mill, $NO_3^-$, $SO_4^{2-}$, and coke plant. These factors generally agreed with those derived by Pekney et al.[65] The coke plant, which was characterized by Cr and copper, may be equivalent to "specialty steel." The CPF and NPR analyses conducted with this study, however, were not very useful, because prevailing southwest to southeast winds dominated throughout the study period.

## Use of Size Information

UNMIX was applied to particle size distributions by Kim et al.[95] Zhou et al.[91] attempted to analyze particle size distribution data with PMF. It is assumed that particles from different sources may have characteristic size distribution profiles (i.e., using $i$ in eq 2 to represent size instead of chemical profiles). Particle size data was selected to exclude periods of particle growth. Five factors were found at Pittsburgh: (1) secondary + primary combustion aerosol; (2) stationary combustion sources; (3) remote traffic; (4) local traffic; and (5) local particle nucleation. These factors were identified by associations with temperature, trace gas pollutants, and transport direction. The first two factors accounted for more than 90% of the particle volume. However, it was difficult to link the factors to major sources resolved by PMF applied to chemical concentrations, such as $NO_3^-$, $SO_4^{2-}$, fugitive dust, and industrial source types. Owing to a lack of size specific source profiles, it is uncertain that the two approaches yielded "the same" factors. Zhou et al.[91] conclude that the "size" factor analysis, by itself, was limited when compared with "standard" receptor modeling based on chemical speciation.

It may be useful to combine chemical composition and size distribution data in receptor models, as indicated in eq 1. Ogulei et al.[96] demonstrated how this could be done in two ways. First, particle concentrations in various size intervals (bins) and species concentrations were pooled as PMF input variables. Twelve factors resolved for Baltimore $PM_{2.5}$ contained not only fractional abundances of chemical species, but also the number of particles in each size bin per unit $PM_{2.5}$ mass emitted. The $NO_3^-$ factor displayed a larger mean particle size than did the $SO_4^{2-}$ factor, although it is uncertain whether $NO_3^-$ is associated with sea salt or dust (neither Na nor Ca were included in the solution). Ultrafine particles were associated with gasoline and diesel exhaust, but not vegetative burning. It appears that particle size distributions can help to identify sources. The second approach involved the PLS solution, which sought to identify common factors in time series of chemical concentrations and particle size distributions. Ogulei et al.[96] identified four factors (latent variables) but the physical significance of those factors was unclear. PLS can be used to determine the degree of dependence between two variables (e.g., size, chemical composition, or meteorology) and whether they can be combined into a receptor model analysis.

## New Receptor Model Development

The PDRM is a hybrid source and receptor solution. A Gaussian dispersion model calculates the pollutant concentrations at receptors based on the emission rates and stack configurations of upwind stationary sources. PDRM uses the results from this model as a constraint to the factor analysis (e.g., $T_{ijklmn}$ in eq 1 may be zero for some sources that are not in sector $n$). PDRM is most suitable for analyzing pollutants dominated by local stationary sources with high-time resolution (subhourly) measurements compatible with the fast-changing meteorology. Park et al.[97] used the PDRM to attribute ambient $SO_2$ at Sydney, FL, to six stationary sources (i.e., two coal-fired and two oil-fired utility plants, one fertilizer, and one incinerator), and estimated $SO_2$ concentration and source emission rates within 10% of measured values. Agreements were not as good for some elements, such as Aluminum (Al), Zn, and Pb, possibly because of other small emitters or varying background levels. Application of PDRM to $SO_2$ and elemental concentrations at Pittsburgh showed

similar performance.[98] This application illustrates that direct plume impacts contribute negligible amounts to $PM_{2.5}$, which was probably not the case before substantial efforts were invested in cleaning Pittsburgh's air.[99] PDRM demonstrates the limitations of receptor models applied to integrated 24-hr measurements. The average wind speed or mixing depth would not be correlated well with the average SCEs. Emission rates may show diurnal and daily variability as well.

### Quantification of SOA

Attempts were made to quantify SOA contributions. Pandis et al.[100] noted that "Primary carbonaceous aerosols dominate during the winter in the east, while SOA is an important contributor during the summer (approaching 40% of the OC) with higher contributions observed during $O_3$ episodes (approaching 80% of the OC)." However, SOA is not always resolved by UNMIX, PMF, or ME solutions to eq 1 or 2. Most input data consist of elements, ions, and C that are not specific to SOA end products.

SOA was estimated from the difference between the measured OC and the sum of OC from identified sources. Schauer and Cass[101] found that an effective variance CMB solution with seven sources (vegetative burning, gasoline, diesel, meat cooking, road dust, natural gas combustion, and vegetative detritus) explained approximately 80% of OC at Fresno during a winter episode but not more than 25% of OC at the Kern Wildlife Refuge, a rural site south of Fresno. The unexplained OC was classified as SOA, although there may have been changes in the source profiles with aging and transport. Alternatively, SOA may be represented by a "single constituent source profile" containing only OC.[35,102,103] It accounts for OC not associated with any other sources and leads to results similar to those of Schauer and Cass.[101]

A tracer solution CMB was tested by Gray et al.[104] using EC as the marker for primary source types.[105] Assuming that primary OC and EC concentrations are correlated at a specific location because they have common sources, SOA is estimated as:

$$OC_{pri} = Z + (OC/EC)_{pri} \times EC \qquad (23)$$

$$OC_{sec} = OC_{tot} - OC_{pri} \qquad (24)$$

Because both OC and EC originate from multiple sources, $(OC/EC)_{pri}$ represents an average OC/EC ratio (i.e., enrichment factor) that varies in time and space. $Z$ is sometimes referred to as "background" OC, which may come from noncombustion sources.[106] In some studies, $(OC/EC)_{pri}$ was determined from the regression slope of the lowest 5–10% of OC and EC concentrations.[107,108] This approach may be appropriate for large datasets in which SOA concentrations are nearly zero in some of the samples. Russell and Allen[106] used independent criteria to select samples to derive a representative $(OC/EC)_{pri}$ ratio for the Houston area. They first used $O_3$ concentration as an indicator of SOA formation and used only daily data with a maximum $O_3$ concentration

less than 40 ppb. Second, they excluded samples containing high concentrations of K and $NO_x$ on the basis that these were contaminated by extreme events such as vegetative burning and $O_3$ scavenging. Data with irregular OC/EC ratios were also eliminated.

For Pittsburgh, Cabada et al.[48] estimated $(OC/EC)_{pri}$ from a local emission inventory for 1995. On the basis of the monthly total OC and EC emissions, they estimated $(OC/EC)_{pri}$ ranging from approximately 1.1 in summer to more than 2 in winter, reflecting the increase of vegetative burning during colder months. Two empirical adjustments were made in the use of eqs 23 and 24: (1) conversion of OC and EC concentrations determined with the thermal/optical reflectance (TOR)[53,54] method to equivalent OC and EC for the National Institute for Occupational Safety and Health (NIOSH) thermal/optical transmittance (TOT) method according to Chow et al.[109] so that they are comparable to source measurements with the NIOSH method; and (2) correction of local OC and EC concentrations with uniform background OC and EC concentrations of 2.9 and 0.5 $\mu g/m^3$, respectively.

Cabada et al.[48] estimated SOA contributions to OC ranging from negligible in winter to approximately 50% in summer. This is consistent with a conceptual model of lower photochemical activity and gaseous precursor concentrations in winter. For the PMF solution,[65] unidentified mass was also much higher in summer than in winter (when it was negligible). Although the question of whether the unidentified mass indicates SOA warrants further investigation, this example illustrates the value of applying multiple source apportionment approaches at the same site.

If emission inventories of SOA precursor gases are available, SOA formation may be estimated using the FAC (Table 1c). Dechapanya et al.[110] used the FAC developed by Grosjean and Seinfeld[111] and Grosjean et al.[112] to estimate SOA production in Houston. Although this model is relatively simple, the use of a constant value for the FAC ignores variability in reactivity under different conditions. One may fine-tune the FAC with a gas/particle partitioning model that considers the total PM mass. Dechapanya et al.[110] compared the SOA and EC production rates, determining an annual average SOA concentration of less than 5 $\mu g/m^3$.

Using primary EC in a CMB tracer solution, Russell and Allen[106] estimated SOA as C concentrations ranging from 0.65 ± 1.11 to 1.15 ± 1.52 $\mu g/m^3$ in Houston, which was approximately 25% of the $PM_{2.5}$ OC. As SOA was not included in the CMB calculations in Houston, calculated mass was lower than the measured mass most of the time.[63]

### Wind and Back Trajectory Analysis

Wind directions and back trajectories are often complements to different receptor model solutions (Table 1d). The advantage is that specific sources within a source type with similar chemical characteristics can be identified. Evaluating conditions along a transport pathway (e.g., cloudy, sunny, and dry) can help to estimate changes in source profiles ($T_{ijklmn}$ in eq 1) during transport. Zhou et al.[46] found the CPF analysis sometimes
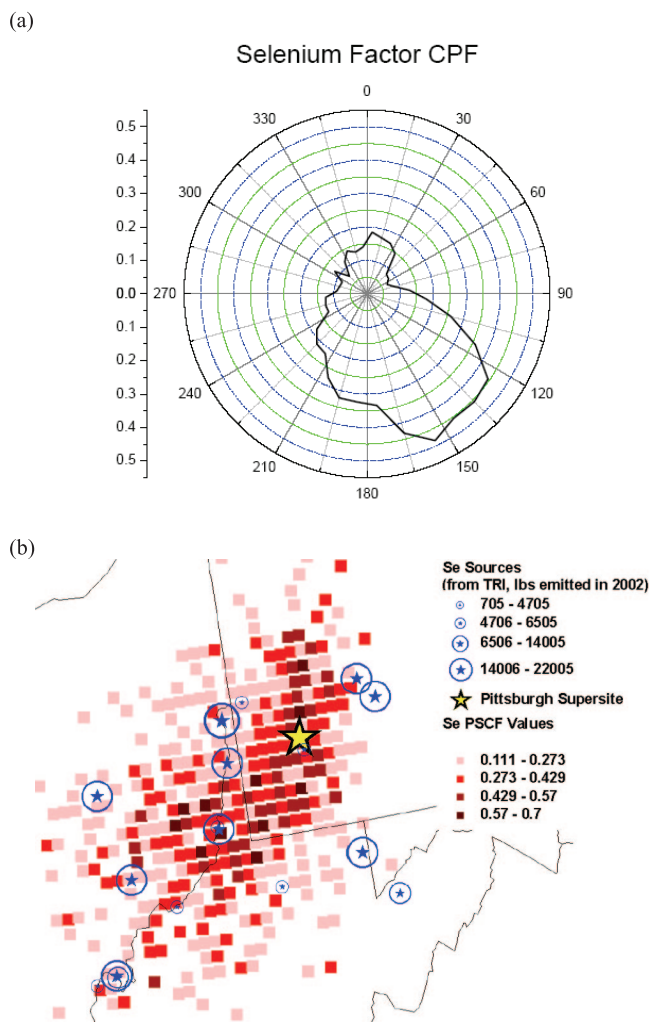
(a)



(b)



**Figure 4.** Comparisons of: (a) CPF and (b) PSCF trajectory models for a Se factor resolved by PMF for the Pittsburg Supersite.[65,113] PSCF shows most probable locations of Se sources southwest of the Supersite, in agreement with the locations of Se sources reported to the Toxics Release Inventory (all coal-fired power plants).

misleading because: (1) there are many directions with very few (or no) wind occurrences; and (2) the distribution of concentrations is far from normal, with a small number of very large values. Figure 4a shows the CPF analysis for a Se factor resolved by PMF at Pittsburgh.[65,113] This factor tends to be associated with southeasterly winds (the dominant wind direction during the study period), whereas there are important Se sources located southwest of the site. In this case, wind speed and boundary layer depth might be more relevant to the pollutant concentration. CPF does not provide a measure of confidence in the probability distribution. NPR does not generate more information than CPF but provides a quantitative estimate of uncertainty. It confirms that many directional signals at this site are associated with uncertainties equal to or greater than their magnitudes, and therefore these signals are not significant.[46]

Back trajectories differ from surface winds in terms of spatial and temporal scales. The back trajectory is probably a better representation of prevailing wind direction if the chosen starting point is above the boundary layer. Back trajectories do not resolve surface wind directions and they often do not represent nighttime transport and mixing when a shallow surface layer often forms. Their uncertainties increase substantially with time.[114,115] The TSA method used to identify source regions for $SO_4^{2-}$ and trace elements in New York state assumed that local source contributions were independent of back trajectories (or prevailing wind directions).[50,116] This assumption contradicts assumptions of CPF and NPR, that is, local source contributions depend on wind patterns. Both can only be correct when there is no correlation between the surface and prevailing wind, and in that case, CPF/NPR only works for local sources and TSA only works for distant sources. In the real world, the accuracy of both approaches is likely site and time dependent. It is worthwhile to evaluate in advance the relationship between surface winds and back trajectories and the extent to which local and regional source contributions depend on them.

There are several variants of TSA, such as PSCF, SQTBA, and RTWC. They attempt to resolve not only the direction of sources from the receptor but also the spatial distribution of source intensities. Pekney et al.[113] compare PSCF with CPF and the spatial distribution of likely sources for Pittsburgh. PSCF and CPF patterns were inconsistent for the sources (e.g., Figure 4). In some cases, sources of these species existed in all directions from the site, resulting in unclear probability fields. It is also likely that the spatial resolution chosen, $0.1° \times 0.1°$, was smaller than the precision of the HYSPLIT model. In a few cases, PSCF and CPF showed reasonable agreement, but the sources could not be identified from the emissions data. This includes emissions enriched in Ga and nickel (Ni).

Zhou et al.[117] used SQTBA and RTWC with the same back trajectories to study $PM_{2.5}$ source locations for two rural sites in New York. Similar patterns were found for some species such as $SO_4^{2-}$, but different patterns were found for others (e.g., geological dust). Compared with PSCF fields that indicate a high source probability when the trajectories pass through a cell, there is no fundamental basis for the SQTBA and RTWC approaches. RTWC redistributes the pollutant concentrations measured at the receptor site along the path of back trajectory according to the residence time, but in reality the residence time is not proportional to the amount of pollutant entrainment into the atmosphere. SQTBA weights each trajectory transport field with pollutant concentration, and this implies that contributions from different trajectory segments are the same. How the SQTBA and RTWC probability fields are related to physical source strength, even qualitatively, is questionable. With the exception of Chen et al.,[118] application of these back trajectory methods ignores the vertical movement of air parcels. Two trajectories with similar horizontal but different vertical motions result from different meteorological patterns will undoubtedly lead

to different ambient pollutant concentrations at a receptor. The effectiveness of current trajectory-based methods must be systematically evaluated.

## CONCLUSIONS

The answer to the question: "How well can we identify and quantify source contributions using receptor models?" is "Reasonably well, but not perfectly." "Reasonably well" means sufficiently to estimate that a contribution is (1) small, (2) about the same as others, or (3) dominant in the sample. This level of accuracy and precision is adequate to make a control decision, although it may not be sufficient to quantify the effects of the control. High-time resolution and chemical specific data from the Supersites Program provided opportunities to apply different solutions to the receptor model equations. The application of multiple solutions sometimes yields consistent source types and SCEs, but this is often not the case. Multiple solutions to the CMB equations should be applied to challenge the results from a single model. The multivariate solutions, such as UNMIX, PMF, and ME, require many samples with different particle size, chemical, temporal, and spatial characteristics, and these types of data are just not available, even from the existing Supersites.

Methods have also been demonstrated to evaluate the stability of receptor models with respect to samples included, species selection, factor and profile selection, and adjustment of parameters in the modeling software. There are too few systematic studies, however, that show how SCEs are affected by changes in these variables.

As with source modeling, receptor modeling contains a subjective element, relying on the modeler's experience and judgment, as well as the available input data. A conceptual model should be established to guide the modeling effort. PDRM represents an important improvement to the factor analysis. It exploits the source model to further constrain the source contributions. More complex models could be used in the same framework.

CMB tracer solutions were commonly used. However, there is rarely a unique tracer for any source. Even if there is, the degree of enrichment of the tracer in the source emission often varies, making it uncertain to estimate PM or VOC contributions using the tracer. It has been shown in the Supersites Program that $(OC/EC)_{pri}$ ratios used for estimating the SOA fraction in OC vary among different sites and seasons.[48,106] Tracer-based method should be considered mostly qualitative, used to provide insights into whether certain sources should be considered in the receptor modeling.

Wind direction and back trajectory analyses have been used to infer source locations and/or spatial distributions. These approaches may be misleading because it is unclear how contributions from local and distant sources change with wind direction and back trajectory. The probability fields generated from ensemble back trajectory analysis (PSCF, SQTBA, and RTWC) do not directly correspond to the source strength. Inconsistencies were observed between different trajectory methods. There is a need to evaluate the effectiveness of these analyses under different atmospheric

conditions; this may also be achieved by a cross comparison between source- and receptor-oriented modeling.

## KNOWLEDGE GAPS AND RECOMMENDATIONS

The approach and benefit of coupling source modeling with receptor modeling needs to be further examined. PDRM shows a good example, but whether this concept can be extrapolated to a more general modeling scheme that includes long-range transport, area sources, and aerosol size and chemistry, is still a question.

### Recommendations

(1) Further compare the PDRM results with effective variance CMB solutions, UNMIX, and PMF using existing Supersite data and high time-resolution data from other monitoring networks.

(2) Evaluate the performance of large-scale receptor models against chemical transport models (CTMs) with respect to source apportionment for both simulated and measured data. Examine consistency for different time resolutions.

(3) Test the response of receptor models with and without constraints provided by the CTMs.

Organic markers were shown to be valuable for source apportionment, but they are usually measured for long-term averages because of high cost and complicated analytical procedures. The diurnal and daily variations of these organic marker concentrations are unknown, and this prevents the inclusion of these species in UNMIX and PMF solutions.

### Recommendations

(1) Develop cost-effective and high time resolution measurements for organic markers. This may be based on the thermal-desorption-gas chromatography/mass spectrometry technique.[61,88,119,120]

(2) Design field experiments to evaluate this technique and acquire sufficient dataset with inorganic (elements and ions) and organic (thermal C fractions and organic species) concentrations at the same temporal resolution.

(3) Test the performance of different models, including PDRM, with and without the C fractions and organic markers.

Efforts should be made to identify profiles that relate sources to ambient pollutants. Current speciation only explains 20–50% of organic mass. The remaining mass could be highly polar humic-like substance (HULIS), which are not fully characterized and quantified.[105] EC is a useful marker for combustion emission, but the differences between EC produced from various combustion sources are uncertain. Wood smoke and diesel exhaust are reported to have different spectral absorption efficiencies and Raman spectra. There were also many single particle measurements made during the Supersites Program,[88,89] but the information is not integrated into receptor models.

### Recommendations

(1) Improve characterization of organic matter, particularly the water-soluble fraction, for speciation

and bulk properties.[88] Understand their abundances in sources and various ambient environments.

(2) Improve the characterization of EC in terms of thermal and optical properties as well as microscopic properties such as size and morphology. Differentiate EC from various combustion sources.

(3) Adapt new methods, including aerosol mass spectroscopy, to receptor models and test their usefulness. This should start with identifying patterns in the source and ambient samples.

## ACKNOWLEDGMENTS

## REFERENCES

1. Zannetti, P. *Air Quality Modeling: Theories, Methodologies, Computational Techniques, and Available Databases and Software Vol. I Fundamentals*; A&WMA: Pittsburgh, 2003.
2. Zannetti, P. *Air Quality Modeling: Theories, Methodologies, Computational Techniques, and Available Databases and Software Vol. II Advanced Topics*; A&WMA: Pittsburgh, 2005.
3. U.S. Environmental Protection Agency. *Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, PM$_{2.5}$, and Regional Haze Draft 3.2*; U.S. Environmental Protection Agency: Research Triangle Park, NC, 2006; available at *http://www.epa.gov/ttn/scram/guidance/guide/draft_final-pm-O3-RH.pdf* (accessed 2007).
4. Buzcu-Guven, B.; Brown, S.G.; Frankel, A.; Hafner, H.R.; Roberts, P.T. Analysis and Apportionment of Organic Carbon and Fine Particulate Matter Sources at Multiple Sites in the Midwestern United States; *J. Air & Waste Manage. Assoc.* **2007**, *57*, 606-619.
5. Feng, Y.C.; Shi, G.L.; Wu, J.H.; Wang, Y.Q.; Zhu, T.; Dai, S.G.; Pei, Y.Q. Source Analysis of Particulate-Phase Polycyclic Aromatic Hydrocarbons in an Urban Atmosphere of a Northern City in China; *J. Air & Waste Manage. Assoc.* **2007**, *57*, 164-171.
6. Fujita, E.M.; Zielinska, B.; Campbell, D.E.; Arnott, W.P.; Sagebiel, J.C.; Mazzoleni, L.; Chow, J.C.; Gabele, P.A.; Crews, W.; Snow, R.; Clark, N.N.; Wayne, W.S.; Lawson, D.R. Variations in Speciated Emissions from Spark-Ignition and Compression-Ignition Motor Vehicles in California's South Coast Air Basin; *J. Air & Waste Manage. Assoc.* **2007**, *57*, 705-720.
7. Pitchford, M.L.; Schichtel, B.A.; Gebhart, K.A.; Barna, M.G.; Malm, W.C.; Tombach, I.H.; Knipping, E.M. Reconciliation and Interpretation of Big Bend National Park's Particulate Sulfur Source Apportionment: Results from the BRAVO Study, Part II; *J. Air & Waste Manage. Assoc.* **2005**, *55*, 1726-1732.
8. Fujita, E.M.; Croes, B.E.; Bennett, C.L.; Lawson, D.R.; Lurmann, F.W.; Main, H.H. Comparison of Emission Inventory and Ambient Concentration Ratios of CO, NMOG, and NO$_x$ in California's South Coast Air Basin; *J. Air & Waste Manage. Assoc.* **1992**, *42*, 264-276.
9. Brock, C.A.; Eatough, D.J.; Soloman, P.A. Preface to Special Section on Particulate Matter: Atmospheric Sciences, Exposure, and the Fourth Colloquium on Particulate Matter and Human Health; *J. Geophys. Res.* **2004**, *109*, D16, doi: 10.1029/2004JD005040.
10. Chow, J.C.; Watson, J.G.; Chen, L.-W.A. Contemporary Inorganic and Organic Speciated Particulate Matter Source Profiles for Geological Material, Motor Vehicles, Vegetative Burning, Industrial Boilers, and Residential Cooking; Draft Report; Desert Research Institute: Reno, NV, 2004.
11. Chow, J.C.; Solomon, P.A. Introduction: a Special Issue of the Journal of the Air & Waste Management Association on the Particulate Matter Supersites Program and Related Studies; *J. Air & Waste Manage. Assoc.* **2006**, *56*, 369.
12. Geller, M.D.; Solomon, P.A. Special Issue of Aerosol Science and Technology for Particulate Matter Supersites Program and Related Studies—Preface; *Aerosol Sci. Technol.* **2006**, *40*, 735-736.
13. Middlebrook, A.; Turner, J.; Solomon, P.A. Special Issue of Atmospheric Environment for Particulate Matter: Atmospheric Sciences, Exposure, and the Fourth Colloquium on PM and Human Health; *Atmos. Environ.* **2004**, *38*, 5179-5181.
14. Ondov, J.M.; Davidson, C.I.; Solomon, P.A. Preface: Special Issue of Aerosol Science and Technology for Particulate Matter: Atmospheric Sciences, Exposure, and the Fourth Colloquium on PM and Human Health; *Aerosol Sci. Technol.* **2004**, *38* (Suppl. 2), 1-2.
15. Sioutas, C.; Pandis, S.N.; Allen, D.T.; Solomon, P.A. Preface: Special Issue of Atmospheric Environment on Findings from EPA's Particulate Matter Supersites Program; *Atmos. Environ.* **2004**, *38*, 3101-3106.
16. Solomon, P.A.; Cowling, E.B.; Weber, R. Preface to the Special Section: Southern Oxidants Study 1999 Atlanta Supersite Project (SOS3); *J. Geophys. Res.* **2003**, *108*, D7, doi: 10.1029/2003/JD003536.
17. Solomon, P.A.; Allen, D. Preface: Special Issue of Aerosol Science and Technology on Findings from the Fine Particulate Matter Supersites Program; *Aerosol Sci. Technol.* **2004**, *38* (Suppl. 1), 1-4.
18. Stanier, C.O.; Solomon, P.A. Preface to Special Section on Particulate Matter Supersites Program and Related Studies; *J. Geophys. Res.* **2006**, *111*, D10S01, doi: 10.1029/2006/JD007513.
19. Chu, S.H.; Paisie, J.W.; Jang, B.W.L. PM Data Analysis: a Comparison of Two Urban Areas: Fresno and Atlanta; *Atmos. Environ.* **2004**, *38*, 3155-3164.
20. Watson, J.G. Visibility: Science and Regulation; *J. Air & Waste Manage. Assoc.* **2002**, *52*, 628-713.
21. Hansen, D.A.; Edgerton, E.; Hartsell, B.; Jansen, J.; Burge, H.; Koutrakis, P.; Rogers, C.; Suh, H.; Chow, J.C.; Zielinska, B.; McMurry, P.; Mulholland, J.; Russell, A.; Rasmussen, R. Air Quality Measurements for the Aerosol Research and Inhalation Epidemiology Study (AIRES); *J. Air & Waste Manage. Assoc.* **2006**, *56*, 1445-1458.
22. Blifford, I.H.; Meeker, G.O. A Factor Analysis Model of Large Scale Pollution; *Atmos. Environ.* **1967**, *1*, 147-157.
23. Blanchard, C.L. Methods for Attributing Ambient Air Pollutants to Emission Sources; *Annu. Rev. Energy Environ.* **1999**, *24*, 329-365.
24. Brook, J.R.; Vega, E.; Watson, J.G. In *Particulate Matter Science for Policy Makers—a NARSTO Assessment, Part 1*. Hales, J.M., Hidy, G.M., Eds.; Cambridge University: London, U.K., 2004; pp 235-281.
25. Chow, J.C.; Watson, J.G. Review of PM$_{2.5}$ and PM$_{10}$ Apportionment for Fossil Fuel Combustion and Other Sources by the Chemical Mass Balance Receptor Model; *Energy Fuels* **2002**, *16*, 222-260.
26. Gordon, G.E. Receptor Models; *Environ. Sci. Technol.* **1980**, *14*, 792-800.
27. Gordon, G.E. Receptor Models; *Environ. Sci. Technol.* **1988**, *22*, 1132-1142.
28. Henry, R.C.; Lewis, C.W.; Hopke, P.K.; Williamson, H.J. Review of Receptor Model Fundamentals; *Atmos. Environ.* **1984**, *18*, 1507-1515.
29. Henry, R.C. History and Fundamentals of Multivariate Air Quality Receptor Models; *Chemom. Intell. Lab. Sys.* **1997**, *37*, 37-42.
30. Henry, R.C. Multivariate Receptor Models: Current Practice and Future Trends; *Chemom. Intell. Lab. Sys.* **2002**, *60*, 43-48.
31. Hopke, P.K. *Receptor Modeling in Environmental Chemistry*; John Wiley & Sons: New York, 1985.
32. Hopke, P.K. *Receptor Modeling for Air Quality Management*; Elsevier: Amsterdam, 1991; Vol. 7.
33. Javitz, H.S.; Watson, J.G.; Guertin, J.P.; Mueller, P.K. Results of a Receptor Modeling Feasibility Study; *J. Air Poll. Control Assoc.* **1988**, *38*, 661-667.
34. Watson, J.G.; Henry, R.C.; Cooper, J.A.; Macias, E.S. In *Atmospheric Aerosol: Source/Air Quality Relationships*; Macias, E.S., Hopke, P.K., Eds.; American Chemical Society: Washington, DC, 1981; pp 89-106.
35. Watson, J.G. Overview of Receptor Model Principles; *J. Air Poll. Control Assoc.* **1984**, *34*, 619-623.
36. Watson, J.G.; Chow, J.C.; Fujita, E.M. Review of Volatile Organic Compound Source Apportionment by Chemical Mass Balance; *Atmos. Environ.* **2001**, *35*, 1567-1584.
37. Watson, J.G.; Zhu, T.; Chow, J.C.; Engelbrecht, J.P.; Fujita, E.M.; Wilson, W.E. Receptor Modeling Application Framework for Particle Source Apportionment; *Chemosphere* **2002**, *49*, 1093-1136.
38. Watson, J.G.; Chow, J.C. Receptor Models for Air Quality Management; *EM* **2004**, October, 27-36.
39. Watson, J.G.; Chow, J.C. In *Air Quality Modeling—Theories, Methodologies, Computational Techniques, and Available Databases and Software. Vol. II— Advanced Topics*; Zannetti, P., Ed.; A&WMA and the Enviro-Comp Institute: Pittsburgh, 2005; pp 455-501.
40. Watson, J.G.; Chow, J.C. Receptor Models for Source Apportionment of Suspended Particles. In *Introduction to Environmental Forensics*, 2nd

ed.; Murphy, B., Morrison, R., Eds.; Academic Press: New York, 2007; pp 279-316.

41. Chow, J.C.; Engelbrecht, J.P.; Watson, J.G.; Wilson, W.E.; Frank, N.H.; Zhu, T. Designing Monitoring Networks to Represent Outdoor Human Exposure; *Chemosphere* **2002**, *49*, 961-978.

42. Hidy, G.M.; Friedlander, S.K. The Nature of the Los Angeles Aerosol. In *Proceedings of the Second International Clean Air Congress*; Englund, H.M., Beery, W.T., Eds.; Academic: New York, 1971; pp 391-404.

43. Friedlander, S.K. The Characterization of Aerosols Distributed with Respect to Size and Chemical Composition II. Classification and Design of Aerosol Measuring Devices; *Aerosol Sci. Technol.* **1971**, *2*, 331-340.

44. Lewis, C.W.; Norris, G.A.; Conner, T.L.; Henry, R.C. Source Apportionment of Phoenix PM$_{2.5}$ Aerosol with the UNMIX Receptor Model; *J. Air &Waste Manage. Assoc.* **2003**, *53*, 325-338.

45. Chen, L.-W.A.; Watson, J.G.; Chow, J.C.; Magliano, K.L. Quantifying PM$_{2.5}$ Source Contributions for the San Joaquin Valley with Multivariate Receptor Models; *Environ. Sci. Technol.* **2007**, *41*, 2818-2826.

46. Zhou, L.M.; Hopke, P.K.; Paatero, P.; Ondov, J.M.; Pancras, J.P.; Pekney, N.J.; Davidson, C.I. Advanced Factor Analysis for Multiple Time Resolution Aerosol Composition Data; *Atmos. Environ.* **2004**, *38*, 4909-4920.

47. Ogulei, D.; Hopke, P.K.; Zhou, L.M.; Paatero, P.; Park, S.S.; Ondov, J.M. Receptor Modeling for Multiple Time Resolved Species: the Baltimore Supersite; *Atmos. Environ.* **2005**, *39*, 3751-3762.

48. Cabada, J.C.; Pandis, S.N.; Robinson, A.L. Sources of Atmospheric Carbonaceous Particulate Matter in Pittsburgh, Pennsylvania; *J. Air & Waste Manage. Assoc.* **2002**, *52*, 732-741.

49. Lemire, K.R.; Allen, D.T.; Klouda, G.A.; Lewis, C.W. Fine Particulate Matter Source Attribution for Southeast Texas Using 14C/13C Ratios; *J. Geophys. Res.* **2002**, *107*, ACH 3-1–ACH 3-7.

50. Dutkiewicz, V.A.; Qureshi, S.; Husain, L.; Schwab, J.J.; Demerjian, K.L. Elemental Composition of PM$_{2.5}$ Aerosols in Queens, New York: Evaluation of Sources of Fine-Particle Mass; *Atmos. Environ.* **2006**, *40* (Suppl. 2), S374-S359.

51. Reff, A.; Eberly, S.I.; Bhave, P.V. Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods; *J. Air & Waste Manage. Assoc.* **2007**, *57*, 146-154.

52. McMurry, P.H.; Shepherd, M.F.; Vickery, J.S. *Particulate Matter Science for Policy Makers: a NARSTO Assessment*; Cambridge University: Cambridge, U.K., 2004.

53. Chow, J.C.; Watson, J.G.; Pritchett, L.C.; Pierson, W.R.; Frazier, C.A.; Purcell, R.G. The DRI Thermal/Optical Reflectance Carbon Analysis System: Description, Evaluation and Applications in U.S. Air Quality Studies; *Atmos. Environ.* **1993**, *27*, 1185-1201.

54. Chow, J.C.; Watson, J.G.; Chen, L.-W.A.; Chang, M.C.O.; Robinson, N.F.; Trimble, D.; Kohl, S. The IMPROVE_A Temperature Protocol for Thermal/Optical Carbon Analysis: Maintaining Consistency with a Long-Term Database; *J. Air & Waste Manage. Assoc.* **2007**, *57*, 1014-1023.

55. Chow, J.C.; Watson, J.G.; Chen, L.-W.A.; Arnott, W.P.; Moosmüller, H.; Fung, K.K. Equivalence of Elemental Carbon by Thermal/Optical Reflectance and Transmittance with Different Temperature Protocols; *Environ. Sci. Technol.* **2004**, *38*, 4414-4422.

56. Chow, J.C.; Watson, J.G.; Chen, L.W.A.; Paredes-Miranda, G.; Chang, M.-C.O.; Trimble, D.; Fung, K.K.; Zhang, H.; Yu, J.Z. Refining Temperature Measures in Thermal/Optical Carbon Analysis; *Atmos. Chem. Phys.* **2005**, *5*, 2961-2972.

57. Kim, E.; Hopke, P.K. Improving Source Identification of Fine Particles in a Rural Northeastern US Area Utilizing Temperature-Resolved Carbon Fractions; *J. Geophys. Res.* **2004**, *109*, 1-13.

58. Kim, E.; Hopke, P.K.; Edgerton, E.S. Improving Source Identification of Atlanta Aerosol Using Temperature Resolved Carbon Fractions in Positive Matrix Factorization; *Atmos. Environ.* **2004**, *38*, 3349-3362.

59. Lee, D.W.; Hopke, P.K.; Rasmussen, D.H.; Wang, H.C.; Mavliev, R. Comparison of Experimental and Theoretical Heterogeneous Nucleation on Ultrafine Carbon Particles; *J. Phys. Chem. B* **2003**, *107*, 13813-13822.

60. Maykut, N.N.; Lewtas, J.; Kim, E.; Larson, T.V. Source Apportionment of PM$_{2.5}$ at an Urban IMPROVE Site in Seattle, Washington; *Environ. Sci. Technol.* **2003**, *37*, 5135-5142.

61. Chow, J.C.; Yu, J.Z.; Watson, J.G.; Ho, S.S.H.; Bohannan, T.L.; Hays, M.D.; Fung, K.K. The Application of Thermal Methods for Determining Chemical Composition of Carbonaceous Aerosols: a Review; *J. Environ. Sci. Health A* **2007**, *42*, 1521-1541.

62. Watson, J.G.; Chow, J.C.; Lowenthal, D.H.; Pritchett, L.C.; Frazier, C.A.; Neuroth, G.R.; Robbins, R. Differences in the Carbon Composition of Source Profiles for Diesel- and Gasoline-Powered Vehicles; *Atmos. Environ.* **1994**, *28*, 2493-2505.

63. Fraser, M.P.; Yue, Z.W.; Buzcu, B. Source Apportionment of Fine Particulate Matter in Houston, TX, Using Organic Molecular Markers; *Atmos. Environ.* **2003**, *37*, 2117-2123.

64. Chow, J.C.; Watson, J.G.; Lowenthal, D.H.; Chen, L.W.A.; Zielinska, B.; Mazzoleni, L.R.; Magliano, K.L. Evaluation of Organic Markers for Chemical Mass Balance Source Apportionment at the Fresno Supersite; *Atmos. Chem. Phys.* **2007**, *7*, 1741-1754.

65. Pekney, N.J.; Davidson, C.I.; Robinson, A.; Zhou, L.M.; Hopke, P.K.; Eatough, D.J.; Rogge, W.F. Major Source Categories for PM$_{2.5}$ in Pittsburgh Using PMF and UNMIX; *Aerosol Sci. Technol.* **2006**, *40*, 910-924.

66. Zhao, W.; Hopke, P.K.; Karl, T. Source Identification of Volatile Organic Compounds in Houston, Texas; *Environ. Sci. Technol.* **2004**, *38*, 1338-1347.

67. Robinson, A.L.; Subramanian, R.; Donahue, N.M.; Bernardo-Bricker, A.; Rogge, W.F. Source Apportionment of Molecular Markers and Organic Aerosols. 1. Polycyclic Aromatic Hydrocarbons and Methodology for Data Visualization; *Environ. Sci. Technol.* **2006**, *40*, 7803-7810.

68. Robinson, A.L.; Subramanian, R.; Donahue, N.M.; Bernardo-Bricker, A.; Rogge, W.F. Source Apportionment of Molecular Markers and Organic Aerosol. 2. Biomass Smoke; *Environ. Sci. Technol.* **2006**, *40*, 7811-7819.

69. Robinson, A.L.; Subramanian, R.; Donahue, N.M.; Bernardo-Bricker, A.; Rogge, W.F. Source Apportionment of Molecular Markers and Organic Aerosol. 3. Food Cooking Emissions; *Environ. Sci. Technol.* **2006**, *40*, 7820-7827.

70. Chow, J.C. Critical Review: Measurement Methods to Determine Compliance with Ambient Air Quality Standards for Suspended Particles; *J. Air & Waste Manage. Assoc.* **1995**, *45*, 320-382.

71. Kim, E.; Hopke, P.K.; Qin, Y. Estimation of Organic Carbon Blank Values and Error Structures of the Speciation Trends Network Data for Source Apportionment; *J. Air & Waste Manage. Assoc.* **2005**, *55*, 1190-1199.

72. Turpin, B.J.; Huntzicker, J.J.; Hering, S.V. Investigation of Organic Aerosol Sampling Artifacts in the Los Angeles Basin; *Atmos. Environ.* **1994**, *28*, 3061-3071.

73. Chow, J.C.; Watson, J.G.; Lowenthal, D.H.; Chen, L.W.A.; Magliano, K.L. Particulate Carbon Measurements in California's San Joaquin Valley; *Chemosphere* **2006**, *62*, 337-348.

74. Chow, J.C.; Chen, L.-W.A.; Watson, J.G.; Lowenthal, D.H.; Magliano, K.L.; Turkiewicz, K.; Lehrman, D. PM$_{2.5}$ Chemical Composition and Spatiotemporal Variability during the California Regional PM$_{10}$/PM$_{2.5}$ Air Quality Study (CRPAQS); *J. Geophys. Res.* **2006**, *111*, D10S04, doi: 10.1029/2005JD006457.

75. Kim, E.; Hopke, P.K. Comparison between Sample-Species Specific Uncertainties and Estimated Uncertainties for the Source Apportionment of the Speciation Trends Network Data; *Atmos. Environ.* **2007**, *41*, 567-575.

76. Eberly, S.I. *EPA PMF 1.1 User's Guide*; U.S. Environmental Protection Agency: Research Triangle Park, NC, 2005.

77. Henry, R.C. *UNMIX Version 2 Manual*; Ronald C. Henry, Ph.D.: West Hills, CA, 2000; available at *http://www.epa.gov/ttn/amtic/files/ambient/pm25/workshop/unmix2.pdf* (accessed 2007).

78. Kim, E.; Hopke, P.K. Source Identifications of Airborne Fine Particles Using Positive Matrix Factorization and U.S. Environmental Protection Agency Positive Matrix Factorization; *J. Air & Waste Manage. Assoc.* **2007**, *57*, 811-819.

79. Henry, R.C. Current Factor Analysis Receptor Models Are Ill-Posed; *Atmos. Environ.* **1987**, *21*, 1815-1820.

80. Henry, R.C. Dealing with Near Collinearity in Chemical Mass Balance Receptor Models; *Atmos. Environ.* **1992**, *26*, 933-938.

81. Watson, J.G.; Robinson, N.F.; Lewis, C.W.; Coulter, C.T.; Chow, J.C.; Fujita, E.M.; Conner, T.L.; Pace, T.G. *CMB8 Application and Validation Protocol for PM$_{2.5}$ and VOCs*. Report No. 1808.2D1. Prepared for U.S. Environmental Protection Agency, Research Triangle Park, NC, by Desert Research Institute, Reno, NV, 1998.

82. Watson, J.G.; Chow, J.C.; Lowenthal, D.H.; Robinson, N.F.; Cahill, C.F.; Blumenthal, D.L. Simulating Changes in Source Profiles from Coal-Fired Power Stations: Use in Chemical Mass Balance of PM$_{2.5}$ in the Mt. Zirkel Wilderness; *Energy Fuels* **2002**, *16*, 311-324.

83. Watson, J.G.; Chow, J.C.; Lurmann, F.W.; Musarra, S. Ammonium Nitrate, Nitric Acid, and Ammonia Equilibrium in Wintertime Phoenix, Arizona; *J. Air & Waste Manage. Assoc.* **1994**, *44*, 405-412.

84. Technical Support Document for the Final Clean Air Interstate Rule—Air Quality Modeling; U.S. Environmental Protection Agency; Office of Air Quality Planning & Standards: Research Triangle Park, NC, 2005.

85. Vayenas, D.V.; Takahama, S.; Davidson, C.I.; Pandis, S.N. Simulation of the Thermodynamics and Removal Processes in the Sulfate-Ammonia-Nitric Acid System during Winter: Implications for PM$_{2.5}$ Control Strategies; *J. Geophys. Res.* **2005**, *110*, doi: 10.1029/2004JD005038.

86. Alpert, D.J.; Hopke, P.K. A Quantitative Determination of Sources in the Boston Urban Aerosol; *Atmos. Environ.* **1980**, *14*, 1137-1146.

87. Cadle, S.H.; Mulawa, P.A.; Hunsanger, E.C.; Nelson, K.; Ragazzi, R.A.; Barrett, R.; Gallagher, G.L.; Lawson, D.R.; Knapp, K.T.; Snow, R. Composition of Light-Duty Motor Vehicle Exhaust Particulate Matter

in the Denver, Colorado Area; *Environ. Sci. Technol.* **1999**, *33*, 2328-2339.

88. Chow, J.C.; Doraiswamy, P.; Watson, J.G.; Chen, L.-W.A.; Ho, S.S.H.; Sodeman, D.A. Advances in Integrated and Continuous Measurements for Particle Mass and Chemical Composition; *J. Air & Waste Manage. Assoc.* **2008**, *58*, 141-163.

89. Solomon, P.A.; Sioutas, C. Continuous and Semicontinuous Monitoring Techniques for Particulate Matter Mass and Chemical Components: a Synthesis of Findings from EPA's Particulate Matter Supersites Program and Related Studies; *J. Air & Waste Manage. Assoc.* **2008**, *58*, 164-195.

90. Sapkota, A.; Symons, J.M.; Kleissl, J.; Wang, L.; Parlange, M.B.; Ondov, J.M.; Breysse, P.N.; Diette, G.B.; Eggleston, P.A.; Buckley, T.J. Impact of the 2002 Canadian Forest Fires on Particulate Matter Air Quality in Baltimore City; *Environ. Sci. Technol.* **2005**, *39*, 24-32.

91. Zhou, L.M.; Kim, E.; Hopke, P.K.; Stanier, C.; Pandis, S.N. Mining Airborne Particulate Size Distribution Data by Positive Matrix Factorization; *J. Geophys. Res.* **2005**, *110*, D07S19, doi: 10.1029/2004JD004707.

92. Polissar, A.V.; Hopke, P.K.; Paatero, P.; Malm, W.C.; Sisler, J.F. Atmospheric Aerosol over Alaska 2. Elemental Composition and Sources; *J. Geophys. Res.* **1998**, *103*, 19045-19057.

93. Polissar, A.V.; Hopke, P.K.; Poirot, R.L. Atmospheric Aerosol over Vermont: Chemical Composition and Sources; *Environ. Sci. Technol.* **2001**, *35*, 4604-4621.

94. Sheesley, R.J.; Schauer, J.J.; Bean, E.; Kenski, D. Trends in Secondary Organic Aerosol at a Remote Site in Michigan's Upper Peninsula; *Environ. Sci. Technol.* **2004**, *38*, 6491-6500.

95. Kim, E.; Hopke, P.K.; Larson, T.V.; Covert, D.S. Analysis of Ambient Particle Size Distributions Using UNMIX and Positive Matrix Factorization; *Environ. Sci. Technol.* **2004**, *38*, 202-209.

96. Ogulei, D.; Hopke, P.K.; Zhou, L.M.; Pancras, J.P.; Nair, N.; Ondov, J.M. Source Apportionment of Baltimore Aerosol from Combined Size Distribution and Chemical Composition Data; *Atmos. Environ.* **2006**, *40* (Suppl. 2), S396-S410.

97. Park, S.S.; Pancras, J.P.; Ondov, J.M.; Poor, N. A New Pseudodeterministic Multivariate Receptor Model for Individual Source Apportionment Using Highly Time-Resolved Ambient Concentration Measurements; *J. Geophys. Res.* **2005**, *110*, D07S15, 1-14, doi: 10.1029/2004JD004664.

98. Park, S.S.; Pancras, J.P.; Ondov, J.M.; Robinson, A. Application of the Pseudo-Deterministic Receptor Model to Resolve Power Plant Influences on Air Quality in Pittsburgh; *Aerosol Sci. Technol.* 2006, *40*, 883-897.

99. Tarr, J.A. Devastation and Renewal: an Environmental History of Pittsburgh and its Region; University of Pittsburgh: Pittsburgh, 2003.

100. Pandis, S.N.; Solomon, P.A.; Scheffe, R. Preface to Special Section on Particulate Matter Supersites; *J. Geophys. Res.* **2005**, *110*, D07S01, doi: 10.1029/2005JD005983.

101. Schauer, J.J.; Cass, G.R. Source Apportionment of Wintertime Gas-Phase and Particle-Phase Air Pollutants Using Organic Compounds as Tracers; *Environ. Sci. Technol.* **2000**, *34*, 1821-1832.

102. Watson, J.G. Ph.D. Dissertation, Oregon Graduate Center, 1979.

103. Chow, J.C.; Watson, J.G.; Lowenthal, D.H.; Solomon, P.A.; Magliano, K.L.; Ziman, S.D.; Richards, L.W. PM₁₀ Source Apportionment in California's San Joaquin Valley; *Atmos. Environ.* **1992**, *26*, 3335-3354.

104. Gray, H.A.; Cass, G.R.; Huntzicker, J.J.; Heyerdahl, E.K.; Rau, J.A. Characteristics of Atmospheric Organic and Elemental Carbon Particle Concentrations in Los Angeles; *Environ. Sci. Technol.* **1986**, *20*, 580-589.

105. Fine, P.M.; Sioutas, C.; Solomon, P.A. Secondary Particulate Matter in the United States: Insights from the Particulate Matter Supersites Program and Related Studies; *J. Air & Waste Manage. Assoc.* **2008**, *58*, 234-253.

106. Russell, M.; Allen, D.T. Seasonal and Spatial Trends in Primary and Secondary Organic Carbon Concentrations in Southeast Texas; *Atmos. Environ.* **2004**, *38*, 3225-3239.

107. Turpin, B.J.; Huntzicker, J.J. Secondary Formation of Organic Aerosol in the Los Angeles Basin: a Descriptive Analysis of Organic and Elemental Carbon Concentrations; *Atmos. Environ.* **1991**, *25*, 207-215.

108. Lim, H.J.; Turpin, B.J. Origins of Primary and Secondary Organic Aerosol in Atlanta: Results of Time-Resolved Measurements during the Atlanta Supersite Experiment; *Environ. Sci. Technol.* **2002**, *36*, 4489-4496.

109. Chow, J.C.; Watson, J.G.; Crow, D.; Lowenthal, D.H.; Merrifield, T.M. Comparison of IMPROVE and NIOSH Carbon Measurements; *Aerosol Sci. Technol.* **2001**, *34*, 23-34.

110. Dechapanya, W.; Russell, M.; Allen, D.T. Estimates of Anthropogenic Secondary Organic Aerosol Formation in Houston, Texas; *Aerosol Sci. Technol.* **2004**, *38* (Suppl. 1), 156-166.

111. Grosjean, D.; Seinfeld, J.H. Parameterization of the Formation Potential of Secondary Organic Aerosols; *Atmos. Environ.* **1989**, *23*, 1733-1747.

112. Grosjean, D.; Williams, E.L. II; Seinfeld, J.H. Atmospheric Oxidation of Selected Terpenes and Related Carbonyls: Gas-Phase Carbonyl Products; *Environ. Sci. Technol.* **1992**, *26*, 1526-1531.

113. Pekney, N.J.; Davidson, C.I.; Zhou, L.M.; Hopke, P.K. Application of PSCF and CPF to PMF-Modeled Sources of PM₂.₅ in Pittsburgh; *Aerosol Sci. Technol.* **2006**, *40*, 952-961.

114. Draxler, R.R.; Dietz, R.M.; Lagomarsino, R.J.; Start, G. Across North America Tracer Experiment (ANATEX): Sampling and Analysis; *Atmos. Environ.* **1991**, *25*, 2815-2836.

115. Draxler, R.R. The Accuracy of Trajectories during ANATEX Calculated Using Dynamic Model Analysis Versus Rawinsonde Observations; *J. Appl. Meteorol.* **1991**, *30*, 1466-1467.

116. Dutkiewicz, V.A.; Qureshi, S.; Khan, A.R.; Ferraro, V.; Schwab, J.; Demerjian, K.L.; Husain, L. Sources of Fine Particulate Sulfate in New York; *Atmos. Environ.* **2004**, *38*, 3179-3189.

117. Zhou, L.M.; Hopke, P.K.; Liu, W. Comparison of Two Trajectory Based Models for Locating Particle Sources for Two Rural New York Sites; *Atmos. Environ.* **2004**, *38*, 1955-1963.

118. Chen, L.-W.A.; Doddridge, B.G.; Dickerson, R.R.; Chow, J.C.; Henry, R.C. Origins of Fine Aerosol Mass in the Baltimore-Washington Corridor: Implications from Observation, Factor Analysis, and Ensemble Air Parcel Back Trajectories; *Atmos. Environ.* **2002**, *36*, 4541-4554.

119. Ho, S.S.H.; Yu, J.Z. Determination of Airborne Carbonyls: Comparison of a Thermal Desorption/GC Method with the Standard DNPH/HPLC Method; *Environ. Sci. Technol.* **2004**, *38*, 862-870.

120. Williams, B.J.; Goldstein, A.H.; Kreisberg, N.M.; Hering, S.V. An In-Situ Instrument for Speciated Organic Composition of Atmospheric Aerosols: Thermal Desorption Aerosol GC/MS-FID (TAG); *Aerosol Sci. Technol.* **2006**, *40*, 627-638.

121. Watson, J.G.; Robinson, N.F.; Lewis, C.W.; Coulter, C.T.; Chow, J.C.; Fujita, E.M.; Lowenthal, D.H.; Conner, T.L.; Henry, R.C.; Willis, R.D. *Chemical Mass Balance Receptor Model Version 8 (CMB) User's Manual*; Desert Research Institute: Reno, NV, 1997.

122. Watson, J.G.; Cooper, J.A.; Huntzicker, J.J. The Effective Variance Weighting for Least Squares Calculations Applied to the Mass Balance Receptor Model; *Atmos. Environ.* **1984**, *18*, 1347-1355.

123. Paatero, P. Least Squares Formulation of Robust Non-Negative Factor Analysis; *Chemom. Intell. Lab. Sys.* **1997**, *37*, 23-35.

124. Paatero, P.; Hopke, P.K.; Song, X.H.; Ramadan, Z. Understanding and Controlling Rotations in Factor Analytical Models; *Chemom. Intell. Lab. Sys.* **2002**, *60*, 253-264.

125. Paatero, P. The Multilinear Engine—a Table-Driven, Least Squares Program for Solving Multilinear Problems, Including the N-Way Parallel Factor Analysis Model; *J. Comput. Graph. Stats.* **1999**, *8*, 854-888.

126. Henry, R.C. Multivariate Receptor Modeling by N-Dimensional Edge Detection; *Chemom. Intell. Lab. Sys.* **2003**, *65*, 179-189.

127. Henry, R.C.; Park, E.S.; Spiegelman, C.H. Comparing a New Algorithm with the Classic Methods for Estimating the Number of Factors; *Chemom. Intell. Lab. Sys.* **1999**, *48*, 91-97.

128. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression: a Basic Tool of Chemometrics; *Chemom. Intell. Lab. Sys.* **2001**, *58*, 109-130.

129. Dams, R.; Rahn, K.A.; Robbins, J.A.; Nifong, G.D.; Winchester, J.W. Multi-Element Analysis of Air Pollution Particulates by Nondestructive Neutron Activation. In *Proceedings of the Second International Clean Air Congress*; Englund, H.M., Beery, W.T., Eds.; Academic: New York, 1971; pp 509-516.

130. Reimann, C.; de Caritat, P. Intrinsic Flaws of Element Enrichment Factors (EFs) in Environmental Geochemistry; *Environ. Sci. Technol.* **2000**, *34*, 5084-5091.

131. Lawson, C.L.; Hanson, R.J. *Solving Least Squares Problems*; Princeton-Hall: Englewood Cliffs, NJ, 1974

132. Wang, D.; Hopke, P.K. The Use of Constrained Least-Squares to Solve the Chemical Mass Balance Problem; *Atmos. Environ.* **1989**, *23*, 2143-2150.

133. Gan, F.; Hopke, P.K. Data Mining of the Relationship between Volatile Organic Components and Transient High Ozone Formation; *Anal. Chim. Acta* **2003**, *490*, 153-158.

134. Ashbaugh, L.L. A Statistical Trajectory Technique for Determining Air Pollution Source Regions; *J. Air Poll. Control Assoc.* **1983**, *33*, 1096-1098.

135. Ashbaugh, L.L.; Myrup, L.O.; Flocchini, R.G. A Principal Component Analysis of Sulphur Concentrations in the Western United States; *Atmos. Environ.* **1984**, *18*, 783-791.

136. Henry, R.C.; Chang, Y.S.; Spiegelman, C.H. Locating Nearby Sources of Air Pollution by Nonparametric Regression of Atmospheric Concentrations on Wind Direction; *Atmos. Environ.* **2002**, *36*, 2237-2244.

137. Yu, K.N.; Cheung, Y.P.; Cheung, T.; Henry, R.C. Identifying the Impact of Large Urban Airports on Local Air Quality by Nonparametric Regression; *Atmos. Environ.* **2004**, *38*, 4501-4507.

138. Parekh, P.P.; Husain, L. Author's Reply to Discussions: 'Trace Element Concentrations in Summer Aerosols at Rural Sites in New York State and Their Possible Sources'; *Atmos. Environ.* **1982**, *16*, 1279-1280.

139. Draxler, R.R. Hybrid Single-Particle Lagrangian Integrated Trajectories (HY-SPLIT): Model Description; National Oceanic and Atmospheric Administration Technical Memorandum ERL ARL-166; National Oceanic and Atmospheric Administration: Silver Spring, MD, 1988.

140. Hopke, P.K.; Li, C.L.; Ciszek, W.; Landsberger, S. The Use of Bootstrapping to Estimate Conditional Probability Fields for Source Locations of Airborne Pollutants; *Chemom. Intell. Lab. Sys.* **1995**, *30*, 69-79.

141. Liu, W.; Hopke, P.K.; VanCuren, R.A. Origins of Fine Aerosol Mass in the Western United States Using Positive Matrix Factorization; *J. Geophys. Res.* **2003**, *108* (D23), 4716, doi: 10.1029/2006JD007978.

142. Hwang, I.; Hopke, P.K. Estimation of Source Apportionment and Potential Source Locations of $PM_{2.5}$ at a West Coastal IMPROVE Site; *Atmos. Environ.* **2007**, *41*, 506-518.

143. Keeler, G.J.; Samson, P.J. Spatial Representativeness of Trace Element Ratios; *Environ. Sci. Technol.* **1989**, *23*, 1358-1364.

144. Samson, P.J. Ensemble Trajectory Analysis of Summer Sulfate Concentrations in New York State; *Atmos. Environ.* **1978**, *12*, 1889-1893.

145. Samson, P.J. Trajectory Analysis of Summertime Sulfate Concentrations in the Northeastern United States; *J. Appl. Meteorol.* **1981**, *19*, 1382-1394.

146. Stohl, A. Trajectory Statistics: a New Method to Establish Source-Receptor Relationships of Air Pollutants and its Application to the Transport of Particulate Sulfate in Europe; *Atmos. Environ.* **1996**, *30*, 579-587.

147. Cheng, M.-D.; Hopke, P.K.; Zeng, Y.A Receptor-Oriented Methodology for Determining Source Regions of Particulate Sulfate at Doreset, Ontario; *J. Geophys. Res.* **1993**, *98*, 16839-16849.

148. Schauer, J.J. Analysis of Particle-Phase Organic Speciation Data to Assess Sources of PM; University of Wisconsin: Madison, WI, 2005.

149. Fraser, M.P.; Buzcu, B.; Yue, Z.W.; McGaughey, G.R.; Desai, N.R.; Allen, D.T.; Seila, R.L.; Lonneman, W.A.; Harley, R.A. Separation of Fine Particulate Matter Emitted from Gasoline and Diesel Vehicles Using Chemical Mass Balancing Techniques; *Environ. Sci. Technol.* **2003**, *37*, 3904-3909.

150. Li, Z.; Hopke, P.K.; Husain, L.; Qureshi, S.; Dutkiewicz, V.A.; Schwab, J.J.; Drewnick, F.; Demerjian, K.L. Sources of Fine Particle Composition in New York City; *Atmos. Environ.* **2004**, *38*, 6521-6529.

151. Liu, W.; Hopke, P.K.; Han, Y.J.; Yi, S.M.; Holsen, T.M.; Cybart, S.; Kozlowski, K.; Milligan, M. Application of Receptor Modeling to Atmospheric Constituents at Potsdam and Stockton, NY; *Atmos. Environ.* **2003**, *37*, 4997-5007.

152. Lee, J.H.; Hopke, P.K.; Turner, J.R. Source Identification of Airborne $PM_{2.5}$ at the St. Louis-Midwest Supersite; *J. Geophys. Res.* **2006**, *111*, D10, doi: 10.1029/2005JD006329.

153. Lee, J.H.; Hopke, P.K. Apportioning Sources of $PM_{2.5}$ in St. Louis, MO Using Speciation Trends Network Data; *Atmos. Environ.* **2006**, *40* (Suppl. 2), S360-S377.

154. Liu, W.; Wang, Y.H.; Russell, A.; Edgerton, E.S. Enhanced Source Identification of Southeast Aerosols Using Temperature-Resolved Carbon Fractions and Gas Phase Components; *Atmos. Environ.* **2006**, *40* (Suppl. 2), S445-S466.

**About the Authors**

John G. Watson and Judith C. Chow are research professors, L.-W. Antony Chen is an assistant research professor, and Douglas H. Lowenthal is an associate research professor with the Division of Atmospheric Sciences at the Desert Research Institute (DRI) in Reno, NV. Prakash Doraiswamy, formerly a postdoctoral research associate at DRI, is currently a postdoctoral research associate with the Atmospheric Sciences Research Center at the University at Albany at the State University of New York, on assignment to the Bureau of Air Quality Analysis and Research, New York State Department of Environmental Conservation, Albany, NY. Please address correspondence to: John G. Watson, Division of Atmospheric Sciences, Desert Research Institute, 2215 Raggio Parkway, Reno, NV 89512; phone: +1-775-674-7046; fax: +1-775-674-7009; e-mail: john.watson@dri.edu.